

# Univariate Volatility Modeling

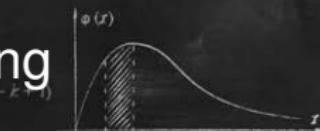
Kevin Sheppard

<http://www.kevinsheppard.com>

Oxford MFE

This version: January 23, 2020

January 2020



$$D_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

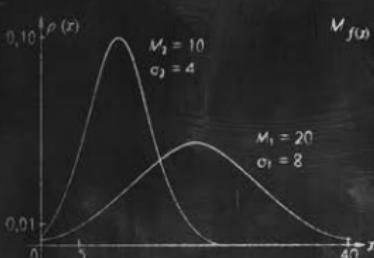
$$M_x = \int_{-\infty}^{+\infty} x \cdot \phi(x) dx$$

$$V_{f(\mu)} = \int_{-\infty}^{+\infty} f(x) \phi(x) dx$$

$$S = v_0 t + \frac{\sigma t^2}{2}$$

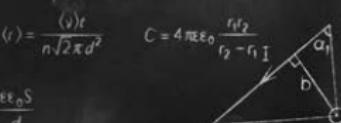
$$F = G \frac{m_1 m_2}{r^2}$$

$$f(v) = 4\pi \left( \frac{m_0}{2\pi k T} \right)^{1/2} v^2 e^{-\frac{mv^2}{2kT}}$$



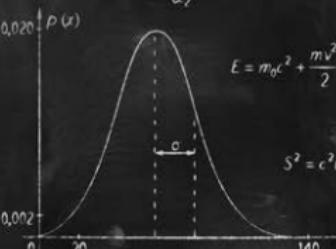
$$\phi(\ln x) d(\ln x) = \frac{1}{\sqrt{2\pi c}} e^{-\frac{(\ln x - a)^2}{2c^2}} d(\ln x) = \frac{1}{\sqrt{2\pi cx}} e^{-\frac{(\ln x - a)^2}{2cx^2}} dx$$

$$B = \frac{\mu_0 I}{2\pi b} (\cos \alpha_1 - \cos \alpha_2)$$



$$A^2 = A_1^2 + A_2^2 + 2 A_1 A_2 \cos(\phi_2 - \phi_1)$$

$$hv = A + \frac{mv^2}{2}$$



$$E = m_0 c^2 + \frac{mv^2}{2}$$

$$m = m_0 / \sqrt{1 - \frac{v^2}{c^2}}$$

$$S^2 = c^2 t^2 - l^2 = i \nu$$

$$r_n = \frac{4\pi \epsilon_0 n^2}{m Z e^2}$$



$$p(A_i | A_F) = p(A_i) p(A_F)$$

$$p(A | B) = \frac{p(AB)}{p(B)}$$

$$p = \lim_{N \rightarrow \infty} \frac{n}{N},$$

$$C = 4\pi \epsilon_0 \frac{r_1 r_2}{r_2 - r_1}$$

$$D_x = \sigma^2 = M_x^2 - (M_x)^2$$

$$p_T(\lambda) = \frac{\lambda^T}{e^T} e^{-\lambda}$$



$$D_x = \sum_{i=1}^k p_i (x_i - M_x)^2$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

$$\phi(v) = 4\sqrt{\frac{k^3}{\pi}} v^2 e^{-kv^2}$$

# Volatility Overview

$$\hat{\sigma}_\theta^2 = \frac{p!}{(n-p)!}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 p(x) dx$$

- What is volatility?
- Why does it change?
- What are ARCH, GARCH, TARCH, EGARCH, SWARCH, ZARCH, APARCH, STARCH, etc. models?
- What does time-varying volatility *look like*?
- What are the basic properties of ARCH and GARCH models?
- What is the news impact curve?
- How are the parameters of ARCH models estimated? What about inference?
- Twists on the standard model
- Forecasting conditional variance
- *Realized Variance*
- Implied Volatility

# What is *volatility*?

$$\hat{\sigma}_B^2 = \frac{\mu!}{(n-k)!}$$



$$\Omega_B = \int_{-\infty}^{+\infty} (x - M_B)^2 p(x) dx$$

- Volatility
- Standard deviation
- Realized Volatility

$$\sigma = \sqrt{T^{-1} \sum_{t=1}^T (r_t - \underline{\mu})^2}$$

- ▶ Other meaning: variance computed from ultra-high frequency (UHF) data

- Conditional Volatility

$$\underline{E_t[\sigma_{t+1}]} = \overline{E[\sigma_{t+1}]}_{\mathcal{E}_C}$$

- Implied Volatility
- Annualized Volatility ( $\sqrt{252} \times$  daily,  $\sqrt{12} \times$  monthly)
- Variance

# Why does volatility change?



$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

- Possible explanations:
  - ▶ News Announcements
  - ▶ Leverage
  - ▶ Volatility Feedback
  - ▶ Illiquidity
  - ▶ State Uncertainty
- None can explain all of the time-variation
- Most theoretical models have none

# This arch?

$$\mu^1 \approx \left[ \frac{1}{\beta} \right]_{\beta=0.8}$$

$$\hat{\sigma}_B^{(t)} = \frac{|\mu|}{(\mu - k)^t}$$



$$Q_2 \leq \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$



No, this arch

$$\mu^1 \approx \left[ \frac{1}{\theta} \right] \sqrt{2 \pi \theta}$$

$$\hat{\sigma}_B^2 = \frac{\mu!}{(\mu - k)!}$$



$$Q_2 \leq \int_{-\infty}^{+\infty} (x - M_2)^2 \phi(x) dx$$



# A basic volatility model: the ARCH(1) model

$$r_t = \epsilon_t$$
$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2$$
$$\epsilon_t = \sigma_t e_t$$
$$\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$
$$\sigma_t^2 = \sigma_t^2$$

- Autoregressive Conditional Heteroskedasticity

- Key model parameters

- ▶  $\omega$  sets the long run level

- ▶  $\alpha$  determines both the persistence and volatility of volatility (VolVol or VolVol)

$$V(x) = E(x) - \underbrace{[E(X^2)]}_{E(X)^2} = M_x^2 - V(x)$$

# Key Properties

- Conditional Mean:  $E_{t-1}[r_t] = E_{t-1}[\epsilon_t] = 0$
- More on this later
  - ▶ Unconditional Mean:  $E[\epsilon_t] = 0$ 
    - Follows directly from the conditional mean and the LIE
- Conditional Variance:  $E_{t-1}[r_t^2] = E_{t-1}[\epsilon_t^2] = \sigma_t^2$
- $\sigma_t^2$  and  $e_t^2$  are independent
- $E_{t-1}[e_t^2] = E[e_t^2] = 1$
- $1 - \alpha_1 > 0$ : Required for stationarity, also  $\alpha_1 \geq 0$ 
  - ▶  $\omega > 0$  is also required for stationarity (technical, but obvious)

$$r_t = M + \Sigma +$$

# Unconditional Variance

- Unconditional Variance

$$E[\varepsilon_t^2] = E[G_t^2] \quad G_t = \varepsilon_t + \bar{\varepsilon}$$

$$E[\varepsilon_t^2] = \frac{\omega}{1 - \alpha_1} = 0.3 \quad E[\varepsilon_t^2] = \frac{w}{\frac{1-\alpha}{1-\alpha}} = \frac{w}{0.3}$$

- Unconditional relates the dynamic parameters to average variance

$$\varepsilon_t = e_t + g_t$$

$$G_{t-1}^2 = w + \alpha \underline{e_{t-1}^2}$$

$$E[G_t^2] = w + \alpha E[G_{t-1}^2]$$

$$\bar{\varepsilon}^2 = w + \alpha \bar{\varepsilon}^2$$

$$\bar{\varepsilon}^2 = \frac{w}{1 - \alpha}$$

$$E[\sigma_t^2] = E[w + \alpha \varepsilon_{t-1}^2] \quad G_t^2 = G_{t-1}^2 + \varepsilon_t^2$$

$$= w + \alpha E[\varepsilon_{t-1}^2]$$

$$= w + \alpha E[e_{t-1}^2 G_{t-1}^2]$$

$$= w + \alpha E[E_{t-2} \left( G_{t-1}^2 + \varepsilon_t^2 \right)]$$

$$= w + \alpha E[G_{t-1}^2 E_{t-2} [e_{t-1}^2]]$$

$$G_{t-1}^2 = \frac{c_{t-1}^2}{c_{t-1}^2 + c_{t-1}^2} = \frac{c_{t-1}^2}{2c_{t-1}^2} = \frac{1}{2}$$

# More properties of the ARCH(1)

- ARCH models are really AR's in disguise
- Add  $\epsilon_t^2 - \sigma_t^2$  to both sides

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\ \sigma_t^2 + \epsilon_t^2 - \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\ \epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\ \epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \nu_t\end{aligned}$$

$\epsilon_t^2 - \sigma_t^2$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \nu_t$$

$\epsilon_t^2 \sigma_t^2 - \sigma_t^2$

$\sigma_t^2 (\epsilon_t^2 - 1)$

- AR(1) in  $\epsilon_t^2$
- $\nu_t = \epsilon_t^2 - \sigma_t^2$  is a mean 0 white noise (WN) process
- $\nu_t$  Captures variance surprise :  $\epsilon_t^2 - \sigma_t^2 = \sigma_t^2 (\epsilon_t^2 - 1)$

# Autocovariance/Autocorrelations

## ■ Autocovariance

$$E[(\epsilon_t^2 - \bar{\sigma}^2)(\epsilon_{t-1}^2 - \bar{\sigma}^2)] = \alpha_1 V[\epsilon_t^2]$$

$$\underline{r_t = c_2}$$

► Same as in AR(1)

■  $j^{th}$  Autocovariance is

■  $j^{th}$  Autocorrelation is

$$\text{Corr}(r_t^L, r_{t-j}^L)$$

$$\text{Corr}(r_t^L, r_{t-j}^L)$$

$$\text{Corr}(\epsilon_t^2, \epsilon_{t-j}^2) = \frac{\alpha_1^j V[\epsilon_t^2]}{V[\epsilon_t^2]} = \alpha_1^j$$

$$\therefore Y_t = \mu_0 + \phi_1 Y_{t-1} + \epsilon_t$$

$$Y_h = \phi_1^h \frac{\mu_0}{1 - \phi_1^h}$$



$$r_t^L = Z_t$$

■ Again, same as AR(1)

■ ARCH(P) is AR(P)

► Just apply results from AR models

# Kurtosis

$$\hat{\sigma}_\mu^2 = \frac{\mu!}{(\mu - k)!}$$



$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - M_x)^2 p(x) dx$$

- Kurtosis Important

- Variance is not constant  $\Rightarrow$  Volatility of Volatility  $> 0$

71

$$\kappa = \frac{E[\epsilon_t^4]}{E[\epsilon_t^2]^2} = \frac{E[\sigma_t^4 e_t^4]}{E[\sigma_t^2 e_t^2]^2} = \frac{E[E_{t+1}[\sigma_{t+1}^4 e_t^4]]}{E[\sigma_t^2]^2} = \frac{3(E[\sigma_t^4])}{E[\sigma_t^2]^2} \geq 3$$

- Alternative:  $E[\sigma_t^4] = V[\sigma_t^2] + E[\sigma_t^2]^2$   
► Law of Iterated Expectations

- In ARCH(1):

$$\kappa = \frac{3(1 - \alpha_1^2)}{(1 - 3\alpha_1^2)} > 3$$

$$6+1 \frac{E_{t+1}[\sigma_{t+1}^4]}{3}$$

$$Y = \frac{\sigma_t^2}{6+1} + \frac{\delta(\sigma_t^2)^2}{6+1}$$

- Finite if  $\alpha_1 < \sqrt{\frac{1}{3}} \approx .577$

$$E[Y^2] = V(Y) + E(Y)^2$$

$$E[(\sigma_t^2)^2] = V(\sigma_t^2) + E(\sigma_t^2)$$

# The ARCH(P) model

## Definition (P<sup>th</sup> Order ARCH)

An Autoregressive Conditional Heteroskedasticity process or order P is given by

$$\begin{aligned} E_{t+1}[r_t] &= E_{t+1}[\mu_t + \epsilon_t] \\ r_t &= \underline{\mu_t} + \epsilon_t \\ \mu_t &= \phi_0 + \phi_1 r_{t-1} + \dots + \phi_s r_{t-s} \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_P \epsilon_{t-P}^2 \\ \epsilon_t &= \sigma_t e_t \\ e_t &\stackrel{\text{i.i.d.}}{\sim} N(0, 1). \end{aligned}$$

$\hat{r}_t = r_t - \underline{\mu_t}$

$\mu_r = \mu$   
 $\mu_r = 0$

- Mean  $\mu_t$  can be an appropriate form - AR, MA, ARMA, ARMAX, etc.
  - ▶  $E_t[r_t - \mu_t] = 0$
- $e_t$  is the standardized residual, often assumed normal
- $\sigma_t^2$  is the conditional variance

$$\begin{aligned} r_t &= \hat{\beta}_0 + \hat{\beta}_1 r_{t-1} \\ \hat{r}_t &= \hat{\beta}_0 + \hat{\beta}_1 r_{t-1} \end{aligned}$$

# Alternative expression of an ARCH(P)

- Model where both mean and variance are time varying
  - ▶ Natural extension of model definition for time varying mean model

$$r_t | \mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t^2)$$

$$\mu_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_s r_{t-s}$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \dots + \alpha_P \epsilon_{t-P}^2$$

$$\epsilon_t = r_t - \mu_t$$

- “ $r_t$  given the information set at time  $t-1$  is conditionally normal with mean  $\mu_t$  and variance  $\sigma_t^2$ ”

# The data

$$\mu \approx \left[ \frac{1}{\sigma} \right] \cdot \text{JSDR}$$

$$\hat{\sigma}_B^2 = \frac{\mu!}{(\mu - k)!}$$



$$P_x = \int_{-\infty}^{+\infty} (x - M_x)^2 p(x) dx$$

## ■ S&P 500

- ▶ Source: Yahoo! Finance
- ▶ Daily January 1, 1999 - December 31, 2018
- ▶ 5,020 observations

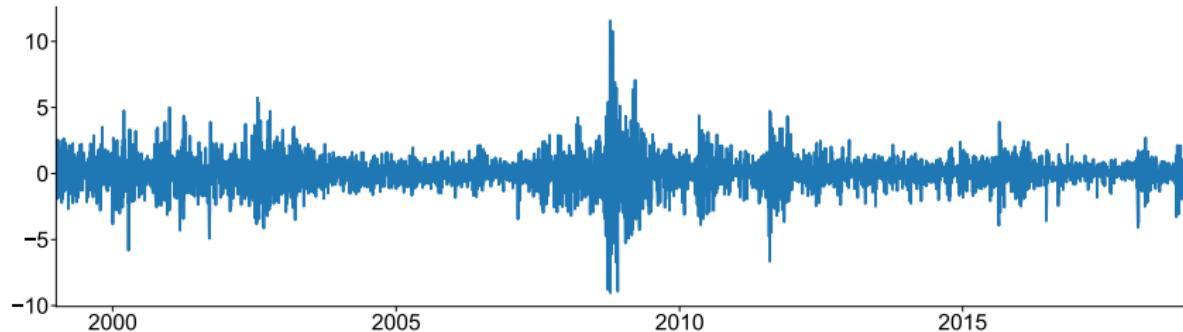
## ■ WTI Spot Prices

- ▶ Source: EIA
- ▶ Daily January 1, 1999 - December 31, 2018
- ▶ 5,007 observations

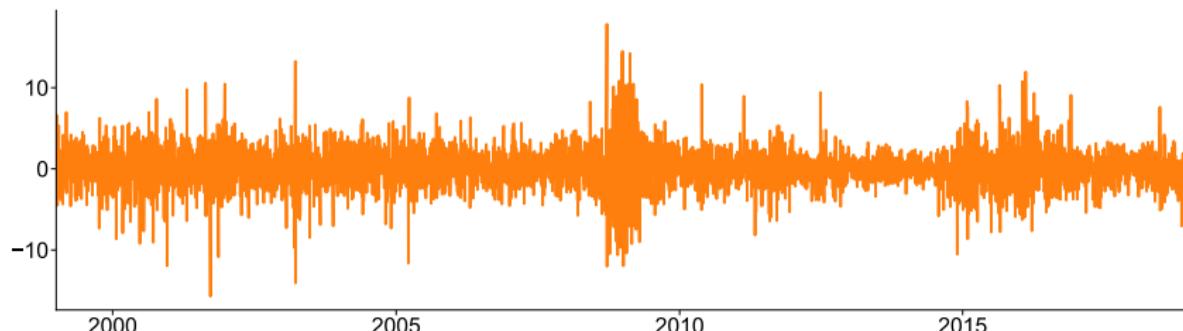
## ■ All represented as 100× returns

# Graphical Evidence of ARCH

S&P 500 Returns

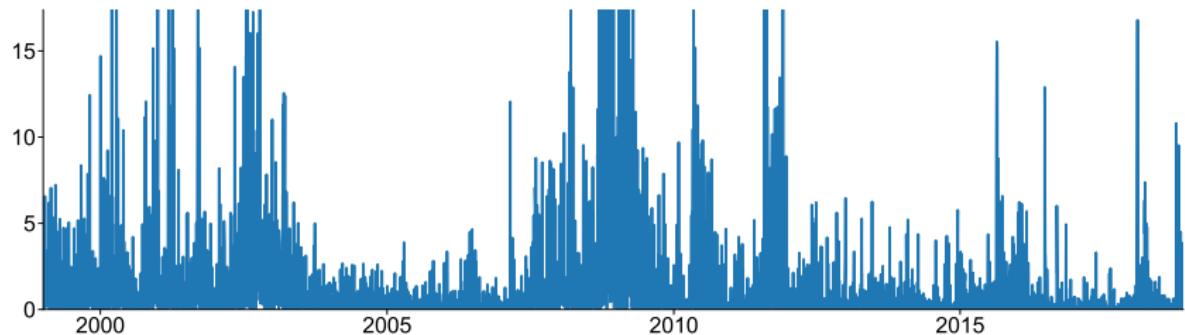


WTI Returns

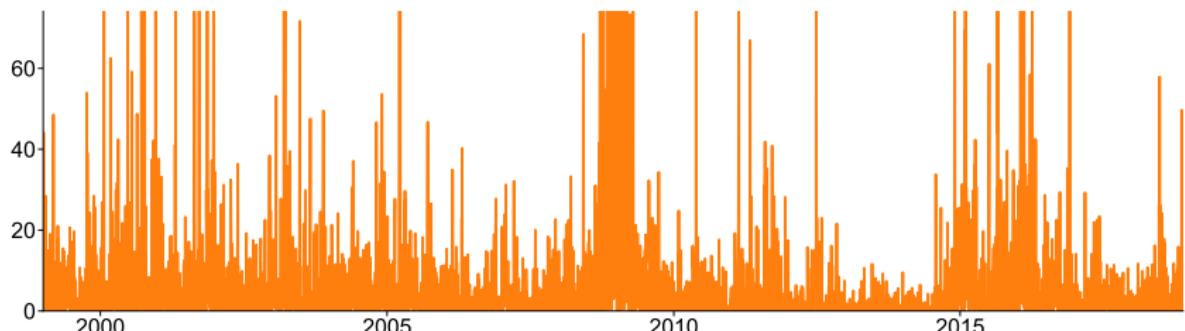


# Graphical Evidence: Squared Data Plot

Squared S&P 500 Returns



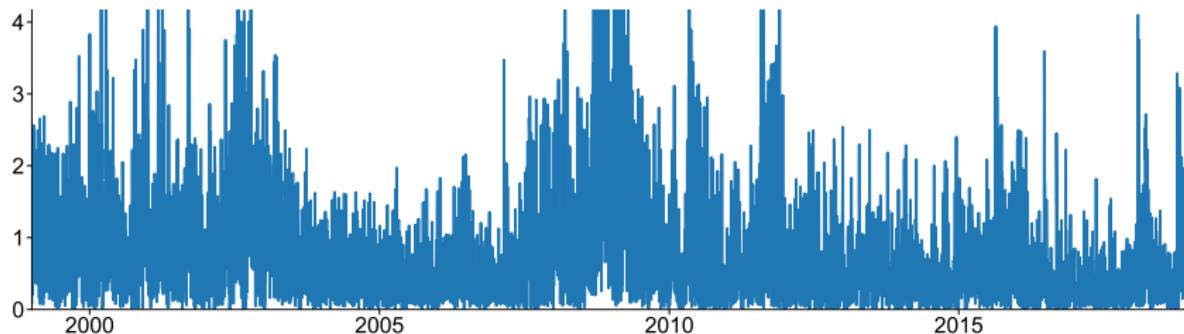
Squared WTI Returns



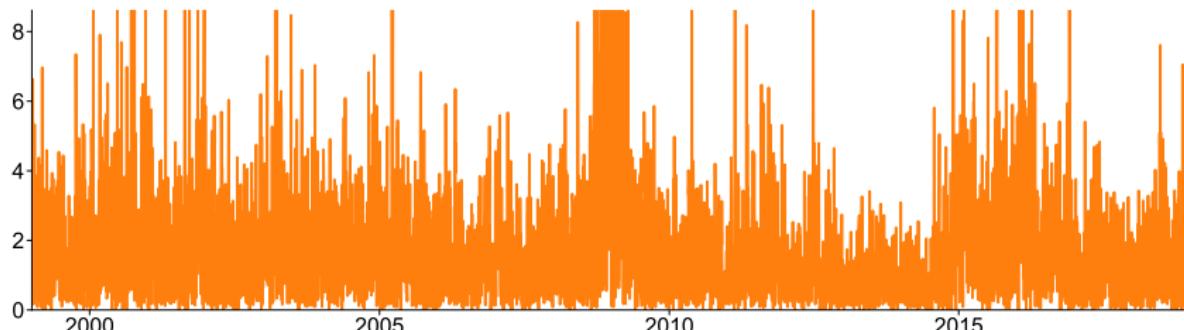
# Graphical Evidence: Absolute Data Plot

$$\sigma_{x,t} = \sqrt{\int_{-\infty}^t (x - M_x)^2 \rho(x) dx}$$

Absolute S&P 500 Returns



Absolute WTI Returns



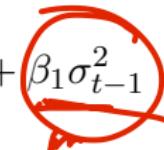
# A simple GARCH(1,1)

$$r_t = \epsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\epsilon_t = \sigma_t e_t$$

$$e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$



- Adds lagged variance to the ARCH model
- ARCH( $\infty$ ) in disguise

$$\sigma_{t-1}^2 = \omega + \alpha \epsilon_{t-2}^2 + \beta \sigma_{t-2}^2$$

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \underline{\beta} (\omega + \alpha \epsilon_{t-2}^2 + \underline{\beta} \sigma_{t-2}^2) \\ &= \omega + \beta \omega + \alpha \epsilon_{t-1}^2 + \beta \alpha \epsilon_{t-2}^2 + \underline{\beta^2 \sigma_{t-2}^2} \\ &= \underline{\omega} + \underline{\beta \omega} + \underline{\beta^2 \sigma_{t-2}^2} + \alpha \epsilon_{t-1}^2 + \beta \alpha \epsilon_{t-2}^2 + \underline{\beta^2 \alpha \epsilon_{t-3}^2} + \beta \alpha \epsilon_{t-3}^2 \\ &= \sum_{j=0}^{\infty} \underline{\beta^j \omega} + \sum_{j=0}^{\infty} \underline{\beta^j} \alpha \epsilon_{t-j-1}^2 + \underline{\beta^3 \sigma_{t-3}^2}\end{aligned}$$

# Important Properties

$$\mu \approx \left[ \frac{1}{n} \sum_{t=1}^n x_t \right]$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

## ■ Unconditional Variance

$$\underline{\sigma^2} = E[\sigma_t^2] = \frac{\omega}{1 - \alpha_1 - \beta_1} = \frac{\underline{\omega}}{\underline{1 - (\alpha_1 + \beta_1)}}$$

## ■ Kurtosis

$$\kappa = \frac{3(1 + \alpha_1 + \beta_1)(1 - \alpha_1 - \beta_1)}{1 - 2\alpha_1\beta_1 - 3\alpha_1^2 - \beta_1^2} > 3$$

## ■ Stationarity

- $\alpha_1 + \beta_1 < 1$
- $\omega > 0, \alpha_1 \geq 0, \beta_1 \geq 0$
- ARMA in disguise

$$\boxed{\epsilon_t^2 - b_t^2}$$

$$\sigma_t^2 + \epsilon_t^2 - \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \epsilon_t^2 - \sigma_t^2$$

$$\epsilon_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \epsilon_t^2 - \sigma_t^2$$

$$\epsilon_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \epsilon_{t-1}^2 - \beta_1 \nu_{t-1} + \nu_t$$

$$\epsilon_t^2 = \omega + (\alpha_1 + \beta_1) \epsilon_{t-1}^2 - \beta_1 \nu_{t-1} + \nu_t$$

$\phi_1$

$-\theta_1$



$$Q_x = \int_{-\infty}^x (x - M_x)^{-\frac{1}{2}} \phi(x) dx$$

.96

.06

.01

w

1 - (α₁ + β₁)

# The Complete GARCH model

## Definition (GARCH(P,Q) process)

A Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process of orders P and Q is defined as

$$r_t = \mu_t + \epsilon_t$$

$$\mu_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_s r_{t-S}$$

$$\sigma_t^2 = \omega + \sum_{p=1}^P \alpha_p \epsilon_{t-p}^2 + \sum_{q=1}^Q \beta_q \sigma_{t-q}^2$$

$$\epsilon_t = \sigma_t e_t$$

$$e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

- Mean model can be altered to fit data –  $AR(S)$  here
- Adds lagged variance to ARCH

# Exponentially Weighted Moving Average Variance

A special case of a GARCH(1,1)

- Restricted model where  $\mu_t = 0$  for all  $t$ ,  $\omega = 0$  and  $\alpha = 1 - \beta$

$$\sigma_t^2 = (1 - \lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2$$

$$\sigma_t^2 = (1 - \lambda) \sum_{i=0}^{\infty} \lambda^i r_{t-i-1}^2$$

- Note that  $\sum_{i=0}^{\infty} \lambda^i = 1/(1-\lambda)$  so that  $(1 - \lambda) \sum_{i=0}^{\infty} \lambda^i = 1$ 
  - ▶ Leads to random-walk-like features

$$n! \approx \left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}$$

$$A_n^k = \frac{n!}{(n-k)!}$$

$$\rho_n = \frac{n!}{(n-n)!} = \frac{n!}{0!}$$

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$$

$$A_n^k = n \cdot (n-1) \cdot \dots \cdot n = n^k$$

$$\tilde{\rho}_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{n_1! n_2! \dots n_k!}$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

$$\tilde{\rho}_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{m!(n-m)!}$$

$$(a+b)^n = C_p^0 a^p + C_p^1 a^{p-1} b^1 + \dots + C_p^{p-1} a^{p-1} b^1 + C_p^p b^n = \sum_{k=0}^p C_p^k a^{p-k} b^k$$

$$p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + p(B|A_3)p(A_3) + \dots + p(B|A_k)p(A_k)$$

$$p(x) = \frac{p(B|A_1)p(A_1)}{p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \dots + p(B|A_k)p(A_k)}$$

$$P_{\mu}(N)$$

# Extensions

$$D_x = M_x^2 - (M_x)^2$$

$$\rho_x(\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \phi(x) dx$$

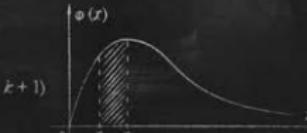
$$M_x = \sum_{i=1}^x P_i X_i$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p = \lim_{N \rightarrow \infty} \frac{f_i}{N}$$

$$C = \frac{PE_0 S}{d}$$

$$D_x = \sum_{i=1}^x P_i (x_i - M_x)^2$$



$$D_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

$$M_x = \int_{-\infty}^{+\infty} x \cdot \phi(x) dx$$

$$V_{f(x)} = \int_{-\infty}^{+\infty} f(x) \phi(x) dx$$

$$S = \eta c^2 + \frac{mc^2}{2}$$

$$F = G \frac{m_1 m_2}{r^2}$$

$$f(x) = \delta g \left( \frac{x_0}{2\pi k T} \right)^N e^{-\frac{m_0 x^2}{2kT}}$$



$$\vec{d} = \frac{m\vec{v}}{2\pi d} (\cos \alpha_1 - \cos \alpha_2)$$



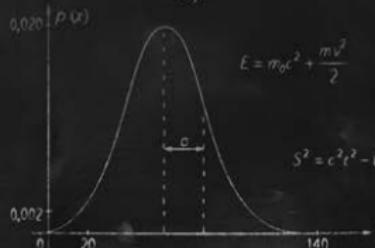
$$d^2 = A_1^2 + A_2^2 + 2 A_1 A_2 \cos(\phi_2 - \phi_1)$$

$$\hbar v = A + \frac{mv^2}{2}$$

$$E = m_0 c^2 + \frac{mv^2}{2}$$

$$m = m_0 / \sqrt{1 - \beta^2}$$

$$S^2 = c^2 t^2 - l^2 = i \hbar v$$



$$r_p = \frac{4\pi \epsilon_0 n^2 r^2}{m Z e^2}$$

# Glosten Jagannathan Runkle-GARCH

- Extends GARCH(1,1) to include an asymmetric term

## Definition (Glosten-Jagannathan-Runkle (GJR)-GARCH process)

A GJR-GARCH(P,O,Q) process is defined as

$$r_t = \mu_t + \epsilon_t$$

$$\mu_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_s r_{t-s}$$

$$\sigma_t^2 = \omega + \sum_{p=1}^P \alpha_p \epsilon_{t-p}^2 + \sum_{o=1}^O \gamma_o \epsilon_{t-o}^2 I_{[\epsilon_{t-o} < 0]} + \sum_{q=1}^Q \beta_q \sigma_{t-q}^2$$

$$\epsilon_t = \sigma_t e_t$$

$$e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

where  $I_{[\epsilon_{t-o} < 0]}$  is an indicator function that takes the value 1 if  $\epsilon_{t-o} < 0$  and 0 otherwise.

# GJR-GARCH(1,1,1) example



$$\sigma_x^2 \approx \int_{-\infty}^{+\infty} (x - M_x)^2 w(x) dx$$

- GJR(1,1,1) model

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2 I_{[\epsilon_{t-1} < 0]} + \beta_1 \sigma_{t-1}^2$$

$$\alpha_1 + \gamma_1 \geq 0$$

$$\alpha_1 \geq 0$$

$$\beta_1 \geq 0$$

$$\omega > 0$$

- $\gamma_1 \epsilon_{t-1}^2 I_{[\epsilon_{t-1} < 0]}$ : Variances are larger after negative shocks than after positive shocks
- “Leverage Effect”

# Threshold ARCH

$$\hat{\sigma}_\theta^2 = \frac{|\mu|}{(n-k)}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

- Threshold ARCH is similar to GJR-GARCH
- Also known as ZARCH (Zakoain (1994)) or AVGARCH when symmetric

## Definition (Threshold ARCH (TARCH) process)

A TARCH(P,O,Q) process is defined

$$r_t = \mu_t + \epsilon_t$$

$$\mu_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_s r_{t-s}$$

$$\sigma_t = \omega + \sum_{p=1}^P \alpha_p |\epsilon_{t-p}| + \sum_{o=1}^O \gamma_o |\epsilon_{t-o}| I_{[\epsilon_{t-o} < 0]} + \sum_{q=1}^Q \beta_q \sigma_{t-q}$$

$$\epsilon_t = \sigma_t e_t$$

$$e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

where  $I_{[\epsilon_{t-o} < 0]}$  is an indicator function that is 1 if  $\epsilon_{t-o} < 0$  and 0 otherwise.

# TARCH(1,1,1) example



- TARCH(1,1,1) model

$$\sigma_t = \omega + \alpha_1 |\epsilon_{t-1}| + \gamma_1 |\epsilon_{t-1}| I_{[\epsilon_{t-1} < 0]} + \beta_1 \sigma_{t-1}$$

$$\alpha_1 + \gamma_1 \geq 0$$

$$\omega > 0, \alpha_1 \geq 0, \beta_1 \geq 0$$

- Note the different power:  $\sigma_t$  and  $|\epsilon_{t-1}|$ 
  - ▶ Model for conditional standard deviation
- Nonlinear variance models complicate some things
  - ▶ Forecasting
  - ▶ Memory of volatility
  - ▶ News impact curves
- GARCH(P,Q) becomes TARCH(P,O,Q) or GJR-GARCH(P,O,Q)
- TARCH and GJR-GARCH are sometimes (*wrongly*) used interchangeably.

# EGARCH

$$\mu \approx \left\{ \begin{array}{l} \frac{\mu_1}{\sigma_1} \\ \frac{\mu_2}{\sigma_2} \end{array} \right\}_{t=0,0}$$

$$\hat{\sigma}_B^2 = \frac{\mu!}{(\mu - k)!}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

## Definition (EGARCH(P,O,Q) process)

An Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) process of order P, O and Q is defined

$$r_t = \mu_t + \epsilon_t$$

$$\mu_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_s r_{t-S}$$

$$\ln(\sigma_t^2) = \omega + \sum_{p=1}^P \alpha_p \left( \left| \frac{\epsilon_{t-p}}{\sigma_{t-p}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{o=1}^O \gamma_o \frac{\epsilon_{t-o}}{\sigma_{t-o}} + \sum_{q=1}^Q \beta_q \ln(\sigma_{t-q}^2)$$

$$\epsilon_t = \sigma_t e_t$$

$$e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

In the original parameterization of Nelson (1991), P and O were required to be identical.

# EGARCH(1,1,1)

$$\mu^t \approx \left\{ \begin{array}{l} \frac{\mu}{\delta} \\ \text{if } \delta \neq 0 \\ \text{else} \end{array} \right.$$

$$\hat{\sigma}_\theta^2 = \frac{\mu^t}{(\mu - \hat{\mu})^2}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

- EGARCH(1,1,1)

$$r_t = \mu + \epsilon_t$$

$$\ln(\sigma_t^2) = \omega + \alpha_1 \left( \left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma_1 \frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta_1 \ln(\sigma_{t-1}^2)$$

$$\epsilon_t = \sigma_t e_t, \quad e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

- Modeling using  $\ln$  removes any parameter restrictions ( $|\beta_1| < 1$ )
- AR(1) with *two* shocks

$$\ln(\sigma_t^2) = \omega + \alpha_1 \left( |e_{t-1}| - \sqrt{\frac{2}{\pi}} \right) + \gamma_1 e_{t-1} + \beta_1 \ln(\sigma_{t-1}^2)$$

- ▶ Symmetric shock  $\left(|e_{t-1}| - \sqrt{\frac{2}{\pi}}\right)$  and asymmetric shock  $e_{t-1}$ 
  - Note, shocks are standardized residuals (unit variance)
- Often provides a better fit than GARCH(P,Q)

# Asymmetric Power ARCH



- Nests ARCH, GARCH, TARCH, GJR-GARCH, EGARCH (almost) and other specifications
- Only present the APARCH(1,1,1):

$$\sigma_t^\delta = \omega + \alpha_1 (|\epsilon_{t-1}| + \gamma_1 \epsilon_{t-1})^\delta + \beta_1 \sigma_{t-1}^\delta$$

$$\alpha_1 > 0, \quad -1 \leq \gamma_1 \leq 1, \quad \delta > 0, \quad \beta_1 \geq 0, \quad \omega > 0$$

- Parameterizes the “power” parameter
- Different values for  $\delta$  affect the persistence.
  - ▶ Lower values  $\Rightarrow$  higher persistence of shocks
    - ARCH:  $\gamma = 0, \beta = 0, \delta = 2$
    - GARCH:  $\gamma = 0, \delta = 2$
    - GJR-GARCH:  $\delta = 2$
    - AVGARCH:  $\gamma = 0, \delta = 1$
    - TARCH:  $\delta = 1$
    - EGARCH: (almost)  $\lim \delta \rightarrow 0$

# S&P Results

$$\mu^j \approx \left[ \frac{1}{\sigma} \right] \int_{-\infty}^{\infty} x \phi(x) dx$$

$$\hat{\rho}_B^{(t)} = \frac{\rho^t}{(\pi - k)^t}$$



$$\Omega_x = \int_{-\infty}^x (x - M_x)^{\alpha} \phi(x) dx$$

## ARCH(5)

$\omega$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	Log Lik.
0.294 (0.000)	0.095 (0.000)	0.204 (0.000)	0.189 (0.000)	0.193 (0.000)	0.143 (0.000)	-7008

## GARCH(1,1)

$\omega$	$\alpha_1$	$\beta_1$	Log Lik.
0.018 (0.000)	0.102 (0.000)	0.885 (0.000)	-6888

## EGARCH(1,1,1)

$\omega$	$\alpha_1$	$\gamma_1$	$\beta_1$	Log Lik.
0.000 (0.909)	0.136 (0.000)	-0.153 (0.000)	0.975 (0.000)	-6767

# Comparing different models



- Comparing models which are not nested can be difficult
- The *News Impact Curve* provides one method
- Defined:

$$n(e_t) = \sigma_{t+1}^2(e_t | \sigma_t^2 = \bar{\sigma}^2)$$

$$NIC(e_t) = n(e_t) - n(0)$$

- Measures the effect of a shock *starting* at the unconditional variance
- Allows for asymmetric shapes

## GARCH(1,1)

$$NIC(e_t) = \alpha_1 \bar{\sigma}^2 e_t^2$$

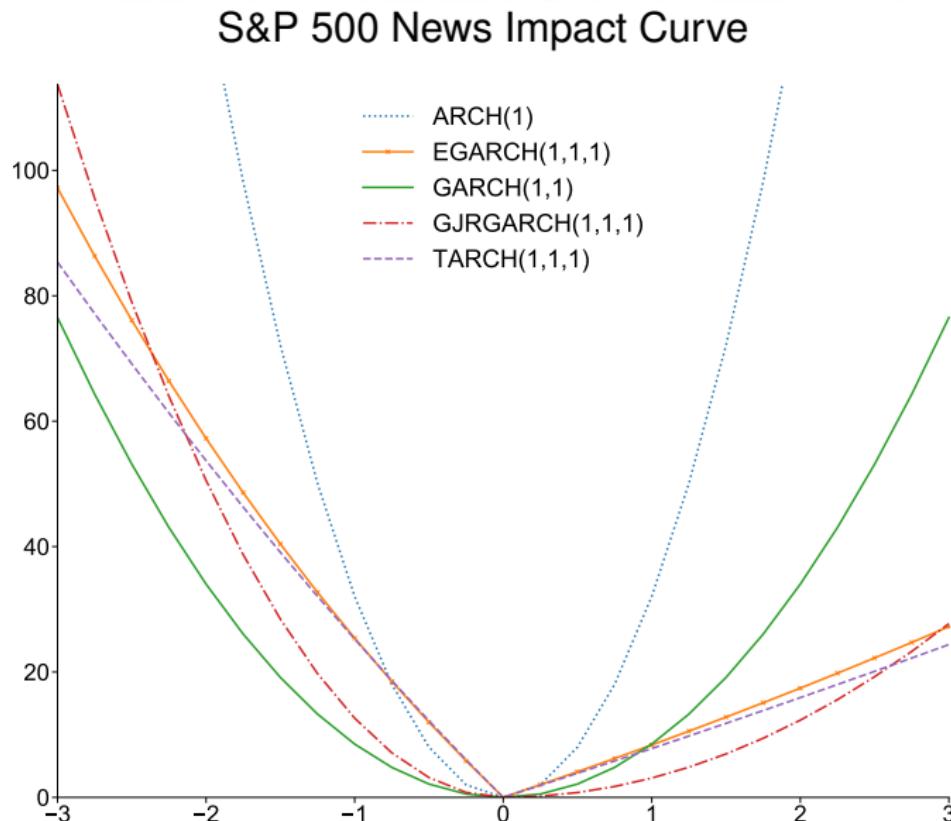
## GJR-GARCH(1,1,1)

$$NIC(e_t) = (\alpha_1 + \gamma_1 I_{[e_t < 0]}) \bar{\sigma}^2 e_t^2$$

## TARCH(1,1,1)

$$NIC(e_t) = (\alpha_1 + \gamma_1 I_{[\epsilon_t < 0]})^2 \bar{\sigma}^2 e_t^2 + (2\omega + 2\beta_1 \bar{\sigma})(\alpha_1 + \gamma_1 I_{[\epsilon_t < 0]}) |\epsilon_t|$$

# S&P 500 News Impact Curves



$$n! \approx \left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}$$

$$A_n^k = \frac{n!}{(n-k)!}$$

$$\rho_n = \frac{n!}{(n-n)!} = \frac{n!}{0!}$$

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$$

$$\lambda_n^k = n \cdot (n-1) \cdot \dots \cdot n = n^k$$

$$\tilde{\rho}_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{n_1! n_2! \dots n_k!}$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

$$\tilde{\rho}_m = \tilde{\rho}_{m, m-1} = \frac{(m+m-1)!}{m!(m-1)!}$$

$$(a+b)^n = C_p^0 a^p + C_p^1 a^{p-1} b^1 + \dots + C_p^{p-1} a^1 b^{p-1} + C_p^p b^p = \sum_{k=0}^p C_p^k a^p - k b^k$$

$$p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + p(B|A_3)p(A_3) + \dots + p(B|A_k)p(A_k)$$

$$\rho(x) = \frac{p(B|A_1)p(A_1)}{p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \dots + p(B|A_k)p(A_k)}$$

$$P_{\mu}(A_i)$$

# Estimation and Inference

$$f_{\theta_0}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta_0)^2}{2}}$$

$$D_x = \hat{M}_x^2 - M_x^2 = (\hat{M}_x^2)^2$$

$$\rho_\varepsilon(x) = \frac{\lambda^\varepsilon}{\varepsilon!} e^{-\lambda^\varepsilon}$$

$$\rho(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \phi(x) dx$$

$$M_x = \sum_{i=1}^k \rho_i x_i$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

$$p = \lim_{N \rightarrow \infty} \frac{f_i}{N}$$

$$C = \frac{PE_n S}{d}$$

$$f_i = \frac{f_i(x)}{\pi \sqrt{d^2 + x^2}}$$

$$C = 4 \cdot \Omega \cdot \pi \cdot \frac{r_1 r_2}{r_2 - r_1}$$

$$\vec{d} = \vec{A}_1^2 + \vec{A}_2^2 + 2 \vec{A}_1 \vec{A}_2 \cos(\phi_2 - \phi_1)$$

$$D_x = \sum_{i=1}^k \rho_i (x_i - M_x)^2$$

$$P(x)$$

$$E = m_0 c^2 + \frac{mv^2}{2}$$

$$m = m_0 / \sqrt{1 - \beta^2}$$

$$S^2 = c^2 t^2 - l^2 = i \nu$$



$$r_n = \frac{4\pi \epsilon_0 n^2 r^2}{m Z e^2}$$

$$D_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

$$M_x = \int_{-\infty}^{+\infty} x \cdot \phi(x) dx$$

$$V_{f(x)} = \int_{-\infty}^{+\infty} f(x) \phi(x) dx$$

$$S = \eta_0^2 \epsilon + \frac{mv^2}{2}$$

$$F = G \frac{m_1 m_2}{r^2}$$

$$f(x) = \delta_E \left( \frac{\pi x}{2 \pi R^2} \right)^N e^{-\frac{m_0 x^2}{2R^2}}$$

# Estimation

$$\mu \approx \left[ \frac{1}{n} \sum_{i=1}^n x_i \right]$$

$$\hat{\sigma}_\theta^2 = \frac{\mu!}{(n-k)!}$$



$$\Omega_x \approx \int_{-\infty}^{+\infty} (x - M_x)^2 p(x) dx$$

$$r_t = \mu_t + \epsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$\epsilon_t = \sigma_t e_t$$

$$e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

■ So:

$$r_t | \mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t^2)$$

- Need initial values for  $\sigma_0^2$  and  $\epsilon_0^2$  to start recursion
  - ▶ Normal Maximum Likelihood is a natural choice

$$f(\mathbf{r}; \boldsymbol{\theta}) = \prod_{t=1}^T (2\pi\sigma_t^2)^{-\frac{1}{2}} \exp\left(-\frac{(r_t - \mu_t)^2}{2\sigma_t^2}\right)$$

$$l(\boldsymbol{\theta}; \mathbf{r}) = \sum_{t=1}^T -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{(r_t - \mu_t)^2}{2\sigma_t^2}.$$

# Inference

$$\mu \approx \left\{ \frac{1}{n} \sum_{i=1}^n x_i \right\}$$

$$\hat{\sigma}_B^2 = \frac{\mu^2}{(n-k)}$$



$$\Omega_2 \approx \int_{-\infty}^{+\infty} (x - M_x)^2 p(x) dx$$

- MLE are asymptotically normal

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}), \quad \mathcal{I} = -E \left[ \frac{\partial^2 l(\theta_0; r_t)}{\partial \theta \partial \theta'} \right]$$

- If data are not conditionally normal, Quasi MLE (QMLE)

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1} \mathcal{J} \mathcal{I}^{-1}), \quad \mathcal{J} = E \left[ \frac{\partial l(\theta_0; r_t)}{\partial \theta} \frac{\partial l(\theta_0; r_t)}{\partial \theta'} \right]$$

- Known as Bollerslev-Wooldridge Covariance estimator in GARCH models
  - ▶ Also known as a “*sandwich*” covariance estimator
  - ▶ Default vcvrobust in MFE GARCH code
  - ▶ White and Newey-West Covariance estimators are also sandwich estimators

# Independence of the mean and variance

The mean and the variance can be estimated consistently using 2-stages. Standard errors are also correct as long as a robust VCV estimator is used.

- Use LS to estimate mean parameters, then use estimated residuals in GARCH
- Efficient estimates one of two ways
- Joint estimation of mean and variance parameters using MLE
- GLS estimation
  - ▶ Estimate mean and variance in 2-steps as above
  - ▶ Re-estimate mean using GLS
  - ▶ Re-estimate variance using new set of residuals

# Independence of the mean and variance

The mean and the variance can be estimated consistently using 2-stages. Standard errors from a robust VCV (White) are also correct.

- First Order Condition:

$$l(r_t | \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{(r_t - \mu_t)^2}{2\sigma_t^2}$$

The second order condition is

$$\frac{\partial^2 l(\mathbf{r} | \theta)}{\partial \mu_t \partial \sigma_t} \frac{\partial \mu_t}{\partial \phi} \frac{\partial \sigma_t^2}{\partial \psi} = \sum_{t=1}^T \frac{(r_t - \mu_t)}{\sigma_t^4} \frac{\partial \mu_t}{\partial \phi} \frac{\partial \sigma_t^2}{\partial \psi}$$

$$E \left[ \frac{\partial^2 l(\mathbf{r} | \theta)}{\partial \mu_t \partial \sigma_t} \frac{\partial \mu_t}{\partial \phi} \frac{\partial \sigma_t^2}{\partial \psi} \right] = E \left[ \sum_{t=1}^T \frac{E_t [r_t - \mu_t]}{\sigma_t^4} \frac{\partial \mu_t}{\partial \phi} \frac{\partial \sigma_t^2}{\partial \psi} \right]$$

$$E \left[ \frac{\partial^2 l(\mathbf{r} | \theta)}{\partial \mu_t \partial \sigma_t} \frac{\partial \mu_t}{\partial \phi} \frac{\partial \sigma_t^2}{\partial \psi} \right] = E \left[ \sum_{t=1}^T \frac{0}{\sigma_t^4} \frac{\partial \mu_t}{\partial \phi} \frac{\partial \sigma_t^2}{\partial \psi} \right]$$

$$E \left[ \frac{\partial^2 l(\mathbf{r} | \theta)}{\partial \mu_t \partial \sigma_t} \frac{\partial \mu_t}{\partial \phi} \frac{\partial \sigma_t^2}{\partial \psi} \right] = 0$$

- Use LS to for mean parameters, then use estimated residuals in GARCH

# GARCH-in-mean models

- Your finance professor would like to believe there is a risk-return tradeoff
- In a GARCH model this can be expressed

$$r_t = \mu + \delta\sigma_t^2 + \epsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$$

$$\epsilon_t = \sigma_t e_t$$

$$e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

- $\delta$  measures the reward per unit variance risk  $\sigma_t^2$ .
- For GIM, ignore my previous rant about estimation
  - ▶ Must be estimated together
- Alternative forms

$$r_t = \mu + \delta\sigma_t + \epsilon_t$$

or

$$r_t = \mu + \delta \ln(\sigma_t^2) + \epsilon_t$$

# GARCH-in-mean on the S&P 500

- Standard GIM model with GARCH(1,1) variance

$$r_t = \mu + \delta g(\sigma_t^2) + \epsilon_t$$

$$\epsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

- Parameter estimates

**S&P 500 Garch-in-Mean Estimates**

	$\mu$	$\delta$	$\omega$	$\alpha$	$\gamma$	$\beta$	Log Lik.
$\sigma^2$	0.004 (0.753)	0.022 (0.074)	0.022 (0.000)	0.000 (0.999)	0.183 (0.000)	0.888 (0.000)	-6773.7
$\sigma$	-0.034 (0.304)	0.070 (0.087)	0.022 (0.000)	0.000 (0.999)	0.182 (0.000)	0.887 (0.000)	-6773.4
$\ln \sigma^2$	0.038 (0.027)	0.030 (0.126)	0.022 (0.000)	0.000 (0.999)	0.183 (0.000)	0.888 (0.000)	-6773.8

# Alternative Distributional Assumptions

- Equity returns are *not* conditionally normal
- Can replace the normal likelihood with a more realistic one
- Common choices:
  - Standardized Student's T
    - ▶ Nests the normal as  $\nu \rightarrow \infty$
  - Generalized error distribution
    - ▶ Nests the normal when  $\nu = 2$
  - Hansen's Skew-T
    - ▶ Captures both skewness and heavy tails
    - ▶ Use *hyperparameters* to control shape ( $\nu$  and  $\lambda$ )
- All can have heavy tails
- Only Skew-T is skewed
- Dozens more in academic research
- But for what gain?

# S&P 500 Density

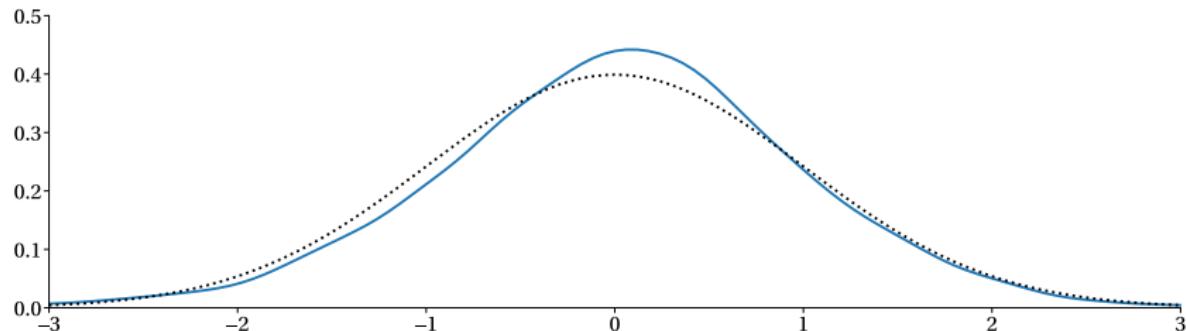
$$\mu \approx \frac{1}{\sigma} \approx 0.8$$

$$\hat{\sigma}_\mu^2 = \frac{\mu^4}{(\mu - k)^2}$$

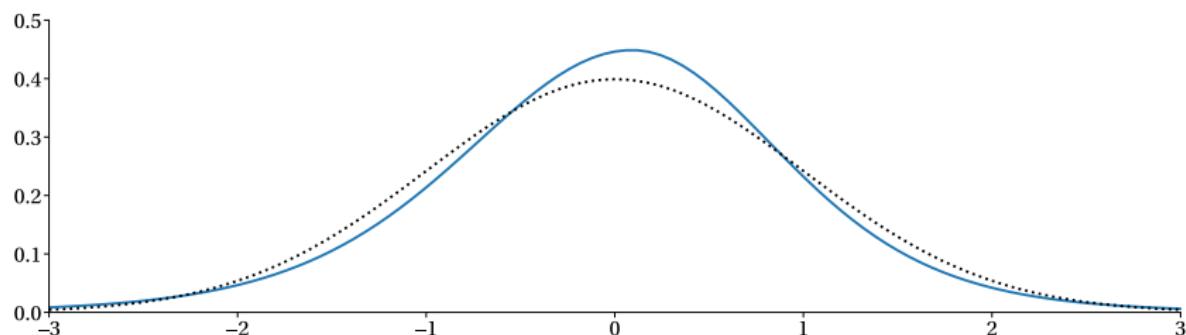


$$\hat{\sigma}_x^2 = \int_{-\infty}^{\infty} (x - M_x)^2 \phi(x) dx$$

Empirical

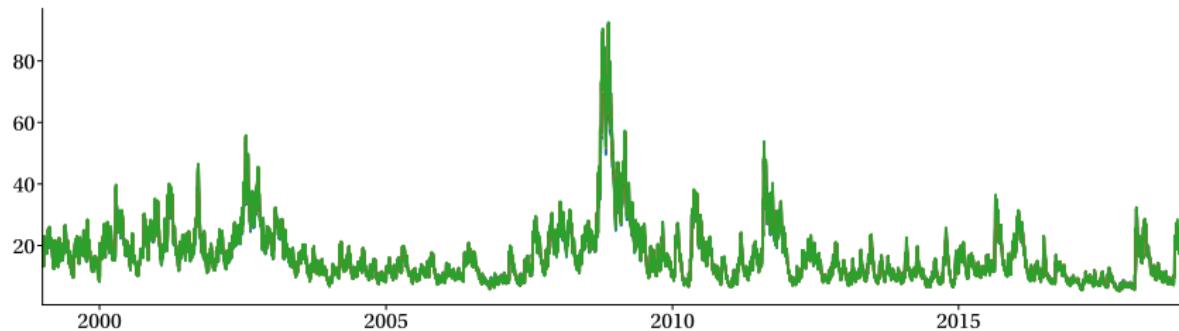


Skew  $t$

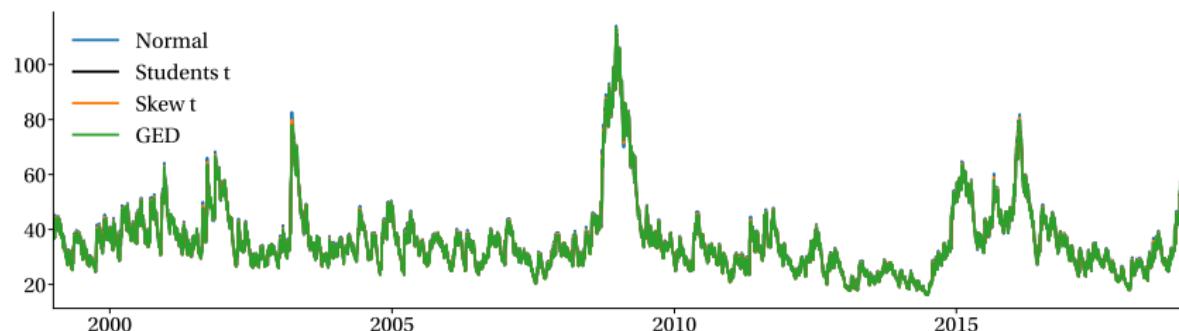


# Effect of dist. choice on estimated volatility

S&P 500



WTI



# What are fat tails?

$$\mu' \approx \left( \frac{1}{\sigma} \right)^2 \cdot \text{ZBB}$$
$$\hat{\sigma}_B^2 = \frac{\mu'}{(\mu - \bar{x})^2}$$

$$\Omega_B = \int_{-\infty}^{+\infty} (x - M_B)^2 \phi(x) dx$$

- “Fat-tailed” and “Thin-tailed”

## Definition (Leptokurtosis)

A random variable  $x_t$  is said to be leptokurtotic if its kurtosis,

$$\kappa = \frac{\text{E}[(x_t - \text{E}[x_t])^4]}{\text{E}[(x_t - \text{E}[x_t])^2]^2}$$

is greater than that of a normal ( $\kappa > 3$ ). Leptokurtotic variables are also known as “heavy tailed” or “fat tailed”.

## Definition (Platykurtosis)

A random variable  $x_t$  is said to be platykurtotic if its kurtosis,

$$\kappa = \frac{\text{E}[(x_t - \text{E}[x_t])^4]}{\text{E}[(x_t - \text{E}[x_t])^2]^2}$$

is less than that of a normal ( $\kappa < 3$ ). Platykurtotic variables are also known as “thin tailed”.

# Model Building

$$\hat{\sigma}_\theta^2 = \frac{p!}{(n-k)!}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

- ARCH and GARCH models are essentially ARMA models
  - ▶ Box-Jenkins Methodology
    - Parsimony principle

Steps:

1. Inspect the ACF and PACF of  $\epsilon_t^2$

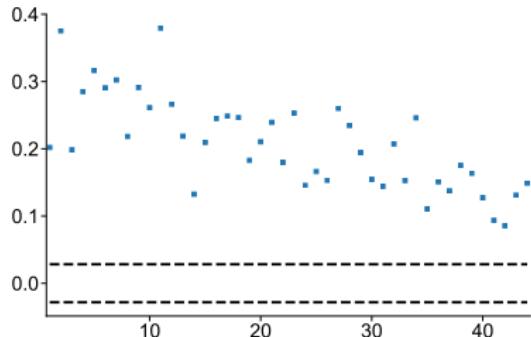
$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 - \beta\nu_{t-1} + \nu_t$$

- ACF indicates  $\alpha$  (or ARCH of any kind)
- PACF indicates  $\beta$

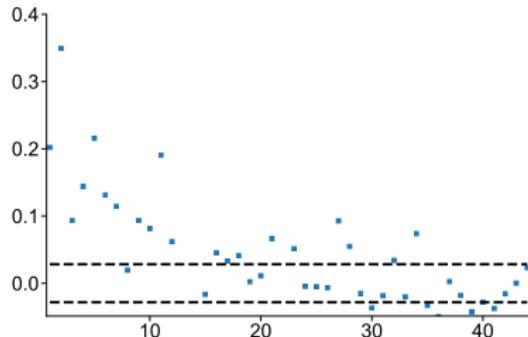
2. Build initial model based on these observation
3. Iterate between model and ACF/PACF of  $\hat{\epsilon}_t^2 = \frac{\epsilon_t^2}{\hat{\sigma}_t^2}$

# S&P 500 $\epsilon_t^2$ ACF/PACF

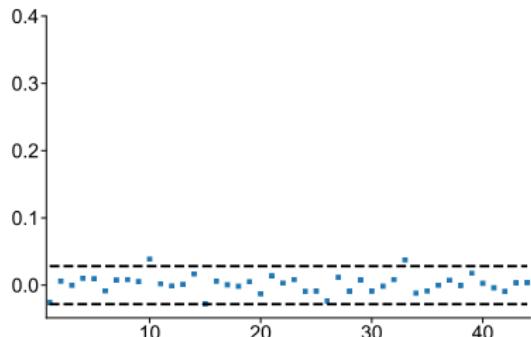
Squared Residuals ACF



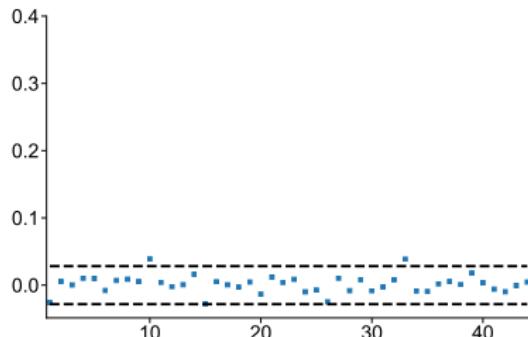
Squared Residuals PACF



Std. Squared Residuals ACF



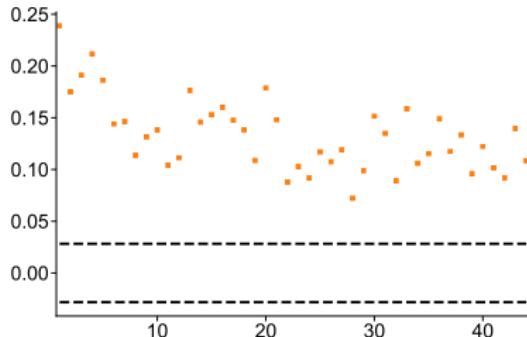
Std. Squared Residuals PACF



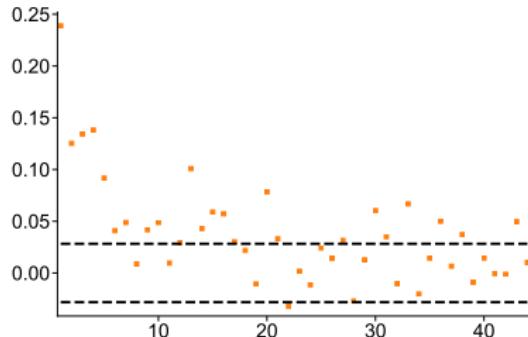
# WTI $\epsilon_t^2$ ACF/PACF

$$\hat{R}_k^2 = \frac{|\rho|}{(n-k)}$$

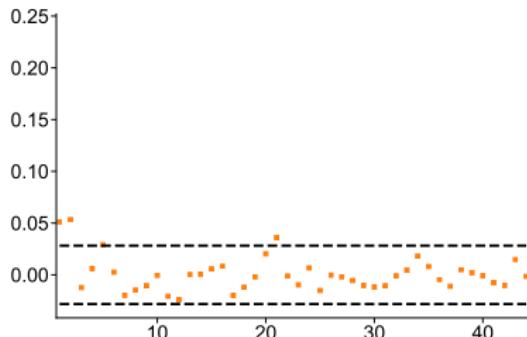
Squared Residuals ACF



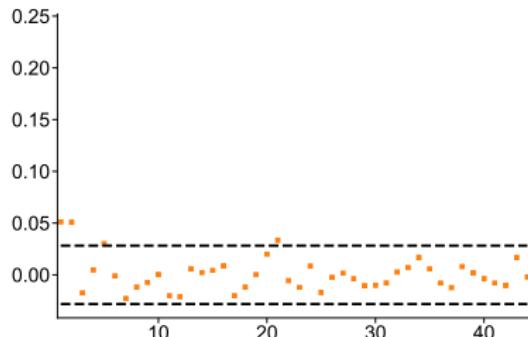
Squared Residuals PACF



Std. Squared Residuals ACF



Std. Squared Residuals PACF



# How I built a model for the S&P 500

	$\alpha_1$	$\alpha_2$	$\gamma_1$	$\gamma_2$	$\beta_1$	$\beta_2$	Log Lik.
GARCH(1,1)	0.102 (0.000)				0.885 (0.000)		-6887.6
GARCH(1,2)	0.102 (0.000)				0.885 (0.000)	0.000 (0.999)	-6887.6
GARCH(2,1)	0.067 (0.003)	0.053 (0.066)			0.864 (0.000)		-6883.5
GJR-GARCH(1,1,1)	0.000 (0.999)		0.185 (0.000)		0.891 (0.000)		-6775.1
GJR-GARCH(1,2,1)	0.000 (0.999)		0.158 (0.000)	0.033 (0.460)	0.887 (0.000)		-6774.5
TARCH(1,1,1)*	0.000 (0.999)		0.172 (0.000)		0.909 (0.000)		-6751.9
TARCH(1,2,1)	0.000 (0.999)		0.165 (0.000)	0.009 (0.786)	0.908 (0.000)		-6751.8
TARCH(2,1,1)	0.000 (0.999)	0.003 (0.936)	0.171 (0.000)		0.907 (0.000)		-6751.9
EGARCH(1,0,1)	0.211 (0.000)				0.979 (0.000)		-6908.4
EGARCH(1,1,1)	0.136 (0.000)		-0.153 (0.000)		0.975 (0.000)		-6766.7
EGARCH(1,2,1)	0.129 (0.000)		-0.213 (0.000)	0.067 (0.045)	0.977 (0.000)		-6761.7
EGARCH(2,1,1)	0.020 (0.651)	0.131 (0.006)	-0.162 (0.000)		0.970 (0.000)		-6757.6

# Testing for (G)ARCH

- ARCH is autocorrelation in  $\epsilon_t^2$
- All ARCH processes have this, whether GARCH or EGARCH or other
  - ▶ ARCH-LM test
  - ▶ Directly test for autocorrelation:

$$\epsilon_t^2 = \phi_0 + \phi_1 \epsilon_{t-1}^2 + \dots + \phi_P \epsilon_{t-P}^2 + \eta_t$$

- ▶  $H_0 : \phi_1 = \phi_2 = \dots = \phi_P = 0$
- ▶  $T \times R^2 \xrightarrow{d} \chi_P^2$
- ▶ Standard LM test from a regression.
- ▶ More powerful test: Fit an ARCH(P) model
- ▶ The forbidden hypothesis

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$H_0 : \alpha_1 = 0, \quad H_1 \alpha > 0$$

$$n! \approx \left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}$$

$$A_n^k = \frac{n!}{(n-k)!}$$

$$\rho_n = \frac{n!}{(n-n)!} = \frac{n!}{0!}$$

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$$

$$\lambda_n^k = n \cdot (n-1) \cdot \dots \cdot n = n^k$$

$$\tilde{\rho}_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{n_1! n_2! \dots n_k!}$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

$$\tilde{\rho}_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{m!(n-m-1)!}$$

$$(d+b)^n = C_p^0 d^n + C_p^1 d^{n-1} b^1 + \dots + C_p^{n-1} d b^{n-1} + C_p^n b^n = \sum_{k=0}^n C_p^k d^{n-k} b^k$$

$$p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + p(B|A_3)p(A_3) + \dots + p(B|A_k)p(A_k)$$

$$\rho(x) = \frac{p(B|A_1)p(A_1)}{p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \dots + p(B|A_k)p(A_k)}$$

$$P_{\mu}(N)$$

# Forecasting

$$D_x = \bar{x}^2 - M_x^2 = (M_x)^2$$

$$\rho_x(\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\rho(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \phi(x) dx$$

$$M_x = \sum_{i=1}^x \rho_i x_i$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p = \lim_{N \rightarrow \infty} \frac{f_i}{N}$$

$$C = \frac{PE_0 S}{d}$$

$$D_x = \sum_{i=1}^x \rho_i (x_i - M_x)^2$$

$$\phi(v) = 4\sqrt{\frac{\epsilon^3}{\pi}} v^2 e^{-v^2}$$



$$D_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

$$M_x = \int_{-\infty}^{+\infty} x \cdot \phi(x) dx$$

$$V_{f(y)} = \int_{-\infty}^{+\infty} f(x) \phi(x) dx$$

$$S = \eta_0^2 c + \frac{m^2}{2}$$

$$F = G \frac{m_1 m_2}{r^2}$$

$$f(x) = \delta g \left( \frac{\pi x}{2 \pi R^2} \right)^N e^{-\frac{m_0^2}{2R^2}}$$



$$\vec{d} = \frac{m_0}{2\pi R} (\cos \alpha_1, \cos \alpha_2)$$



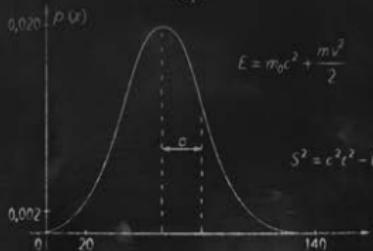
$$d^2 = A_1^2 + A_2^2 + 2 A_1 A_2 \cos(\phi_2 - \phi_1)$$

$$\hbar v = A + \frac{m v^2}{2}$$

$$E = m_0 c^2 + \frac{m v^2}{2}$$

$$m = m_0 / \sqrt{1 - \beta^2}$$

$$S^2 = c^2 t^2 - l^2 = i \hbar v$$



$$r_n = \frac{4\pi \epsilon_0 n^2 n^2}{m Z e^4}$$

# Forecasting: ARCH(1)



## ■ Simple ARCH model

$$\epsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2$$

- ▶ **1-step ahead forecast is known today**
- ▶ All ARCH-family models have this property

$$\epsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2$$

$$\begin{aligned} E_t[\sigma_{t+1}^2] &= E_t[\omega + \alpha_1 \epsilon_t^2] \\ &= \omega + \alpha_1 \epsilon_t^2 \end{aligned}$$

- ▶ **Note:**  $E_t[\epsilon_{t+1}^2] = E_t[e_{t+1}^2 \sigma_{t+1}^2] = \sigma_{t+1}^2 E_t[e_{t+1}^2] = \sigma_{t+1}^2$
- ▶ Further:

$$E_t[\epsilon_{t+h}^2] = E_t[E_{t+h-1}[e_{t+h}^2 \sigma_{t+h}^2]] = E_t[E_{t+h-1}[e_{t+h}^2] \sigma_{t+h}^2] = E_t[\sigma_{t+h}^2]$$

# Forecasting: ARCH(1)



- 2-step ahead

$$\text{E}_t[\sigma_{t+2}^2] =$$

- $h$ -step ahead forecast

$$\text{E}_t[\sigma_{t+h}^2] = \sum_{i=0}^{h-1} \alpha_1^i \omega + \alpha_1^h \epsilon_t^2$$

- ▶ Just the AR(1) forecasting formula
  - Why?

# Forecasting: GARCH(1,1)



- 1-step ahead

$$\begin{aligned}\text{E}_t[\sigma_{t+1}^2] &= \text{E}_t[\omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2] \\ &= \omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2\end{aligned}$$

- 2-step ahead

$$\text{E}_t[\sigma_{t+2}^2] =$$

# Forecasting: GARCH(1,1)



- $h$ -step ahead

$$\text{E}_t[\sigma_{t+h}^2] = \sum_{i=0}^{h-1} (\alpha_1 + \beta_1)^i \omega + (\alpha_1 + \beta_1)^{h-1} (\alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2)$$

- Also essentially an AR(1), technically ARMA(1,1)

# Forecasting: TARCH(1,0,0)



$$\Omega_x = \int_{-\infty}^{\infty} (x - M_x)^2 \omega(x) dx$$

- This one is a mess

- ▶ *Nonlinearities* cause problems

- All ARCH-family models are nonlinear, but some are **linearity** in  $\epsilon_t^2$
    - Others are not

$$\sigma_t = \omega + \alpha_1 |\epsilon_{t-1}|$$

- ▶ **Forecast for  $t+1$  is known at time  $t$** 
    - **Always, always, always, ...**

$$\begin{aligned}\text{E}_t[\sigma_{t+1}^2] &= \text{E}_t[(\omega + \alpha_1 |\epsilon_t|)^2] \\ &= \text{E}_t[\omega^2 + 2\omega\alpha_1 |\epsilon_t| + \alpha_1^2 \epsilon_t^2] \\ &= \omega^2 + 2\omega\alpha_1 \text{E}_t[|\epsilon_t|] + \alpha_1^2 \text{E}_t[\epsilon_t^2] \\ &= \omega^2 + 2\omega\alpha_1 |\epsilon_t| + \alpha_1^2 \epsilon_t^2\end{aligned}$$

# TARCH(1,0,0) continued...



- Multi-step is less straightforward

$$\begin{aligned} E_t[\sigma_{t+2}^2] &= E_t[(\omega + \alpha_1 |\epsilon_{t+1}|)^2] \\ &= E_t[\omega^2 + 2\omega\alpha_1 |\epsilon_{t+1}| + \alpha_1^2 \epsilon_{t+1}^2] \\ &= \omega^2 + 2\omega\alpha_1 E_t[|\epsilon_{t+1}|] + \alpha_1^2 E_t[\epsilon_{t+1}^2] \\ &= \omega^2 + 2\omega\alpha_1 E_t[|e_{t+1}| \sigma_{t+1}] + \alpha_1^2 E_t[e_t^2 \sigma_{t+1}^2] \\ &= \omega^2 + 2\omega\alpha_1 E_t[|e_{t+1}|] E_t[\sigma_{t+1}] + \alpha_1^2 E_t[e_t^2] E_t[\sigma_{t+1}^2] \\ &= \omega^2 + 2\omega\alpha_1 E_t[|e_{t+1}|](\omega + \alpha_1 |\epsilon_t|) + \alpha_1^2 \cdot 1 \cdot (\omega^2 + 2\omega\alpha_1 |\epsilon_t| + \alpha_1^2 \epsilon_t^2) \end{aligned}$$

If  $e_{t+1} \sim N(0, 1)$ ,  $E[|e_{t+1}|] = \sqrt{\frac{2}{\pi}}$

$$E_t[\sigma_{t+2}^2] = \omega^2 + 2\omega\alpha_1 \sqrt{\frac{2}{\pi}}(\omega + \alpha_1 |\epsilon_t|) + \alpha_1^2(\omega^2 + 2\omega\alpha_1 |\epsilon_t| + \alpha_1^2 \epsilon_t^2)$$

# Assessing forecasts: Augmented MZ

- Start from  $E_t[r_{t+h}^2] \approx \sigma_{t+h|t}^2$

- Standard Augmented MZ regression:

$$\epsilon_{t+h}^2 - \hat{\sigma}_{t+h|t}^2 = \gamma_0 + \gamma_1 \hat{\sigma}_{t+h|t}^2 + \gamma_2 z_{1t} + \dots + \gamma_{K+1} z_{Kt} + \eta_t$$

- $\eta_t$  is heteroskedastic in proportion to  $\sigma_t^2$ : Use GLS.
  - An improved GMZ regression (GMZ-GLS)

$$\frac{\epsilon_{t+h}^2 - \hat{\sigma}_{t+h|t}^2}{\hat{\sigma}_{t+h|t}^2} = \gamma_0 \frac{1}{\hat{\sigma}_{t+h|t}^2} + \gamma_1 1 + \gamma_2 \frac{z_{1t}}{\hat{\sigma}_{t+h|t}^2} + \dots + \gamma_{K+1} \frac{z_{Kt}}{\hat{\sigma}_{t+h|t}^2} + \nu_t$$

- Better to use *Realized Variance* to evaluate forecasts

$$RV_{t+h} - \hat{\sigma}_{t+h|t}^2 = \gamma_0 + \gamma_1 \hat{\sigma}_{t+h|t}^2 + \gamma_2 z_{1t} + \dots + \gamma_{K+1} z_{Kt} + \eta_t$$

- Also can use GLS version
  - Both  $RV_{t+h}$  and  $\epsilon_{t+h}^2$  are proxies for the variance at  $t+h$ 
    - RV is just better, often  $10\times+$  more precise

# Assessing forecasts: Diebold-Mariano

- Relative forecast performance

- ▶ MSE loss

$$\delta_t = \left( \epsilon_{t+h}^2 - \hat{\sigma}_{A,t+h|t}^2 \right)^2 - \left( \epsilon_{t+h}^2 - \hat{\sigma}_{B,t+h|t}^2 \right)^2$$

- ▶  $H_0 : E[\delta_t] = 0, H_1^A : E[\delta_t] < 0, H_1^B : E[\delta_t] > 0$

$$\hat{\delta} = R^{-1} \sum_{r=1}^R \delta_r$$

- ▶ Standard t-test, 2-sided alternative
  - ▶ Newey-West covariance always needed
  - ▶ Better DM using “Q-Like” loss (Normal log-likelihood “Kernel”)

$$\delta_t = \left( \ln(\hat{\sigma}_{A,t+h|t}^2) + \frac{\epsilon_{t+h}^2}{\hat{\sigma}_{A,t+h|t}^2} \right) - \left( \ln(\hat{\sigma}_{B,t+h|t}^2) + \frac{\epsilon_{t+h}^2}{\hat{\sigma}_{B,t+h|t}^2} \right)$$

- ▶ Patton & Sheppard (2009)

$$n! \approx \left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}$$

$$A_n^k = \frac{n!}{(n-k)!}$$

$$\rho_n = \frac{n!}{(n-n)!} = \frac{n!}{0!}$$

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$$

$$\lambda_n^k = n \cdot (n-1) \cdot \dots \cdot n = n^k$$

$$\tilde{\rho}_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{n_1! n_2! \dots n_k!}$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

$$\tilde{C}_n^m = \frac{(n+m-1)!}{m!(n-1)!}$$

$$(a+b)^n = C_p^0 a^p + C_p^1 a^{p-1} b^1 + \dots + C_p^{p-1} a^1 b^{p-1} + C_p^p b^n = \sum_{k=0}^n C_p^k a^p - k b^k$$

$$p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + p(B|A_3)p(A_3) + \dots + p(B|A_k)p(A_k)$$

$$\rho(x) = \frac{p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \dots + p(B|A_k)p(A_k)}{p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \dots + p(B|A_k)p(A_k)}$$

$$P_{\mu}(N)$$

$$P_1$$

$$P_2$$

$$P_3$$

$$P_4$$

$$P_5$$

$$P_6$$

$$P_7$$

$$P_8$$

$$P_9$$

$$P_{10}$$

$$P_{11}$$

$$P_{12}$$

$$P_{13}$$

$$P_{14}$$

$$P_{15}$$

$$P_{16}$$

$$P_{17}$$

$$P_{18}$$

$$P_{19}$$

$$P_{20}$$

$$P_{21}$$

$$P_{22}$$

$$P_{23}$$

$$P_{24}$$

$$P_{25}$$

$$P_{26}$$

$$P_{27}$$

$$P_{28}$$

$$P_{29}$$

$$P_{30}$$

$$P_{31}$$

$$P_{32}$$

$$P_{33}$$

$$P_{34}$$

$$P_{35}$$

$$P_{36}$$

$$P_{37}$$

$$P_{38}$$

$$P_{39}$$

$$P_{40}$$

$$P_{41}$$

$$P_{42}$$

$$P_{43}$$

$$P_{44}$$

$$P_{45}$$

$$P_{46}$$

$$P_{47}$$

$$P_{48}$$

$$P_{49}$$

$$P_{50}$$

$$P_{51}$$

$$P_{52}$$

$$P_{53}$$

$$P_{54}$$

$$P_{55}$$

$$P_{56}$$

$$P_{57}$$

$$P_{58}$$

$$P_{59}$$

$$P_{60}$$

$$P_{61}$$

$$P_{62}$$

$$P_{63}$$

$$P_{64}$$

$$P_{65}$$

$$P_{66}$$

$$P_{67}$$

$$P_{68}$$

$$P_{69}$$

$$P_{70}$$

$$P_{71}$$

$$P_{72}$$

$$P_{73}$$

$$P_{74}$$

$$P_{75}$$

$$P_{76}$$

$$P_{77}$$

$$P_{78}$$

$$P_{79}$$

$$P_{80}$$

$$P_{81}$$

$$P_{82}$$

$$P_{83}$$

$$P_{84}$$

$$P_{85}$$

$$P_{86}$$

$$P_{87}$$

$$P_{88}$$

$$P_{89}$$

$$P_{90}$$

$$P_{91}$$

$$P_{92}$$

$$P_{93}$$

$$P_{94}$$

$$P_{95}$$

$$P_{96}$$

$$P_{97}$$

$$P_{98}$$

$$P_{99}$$

$$P_{100}$$

$$P_{101}$$

$$P_{102}$$

$$P_{103}$$

$$P_{104}$$

$$P_{105}$$

$$P_{106}$$

$$P_{107}$$

$$P_{108}$$

$$P_{109}$$

$$P_{110}$$

$$P_{111}$$

$$P_{112}$$

$$P_{113}$$

$$P_{114}$$

$$P_{115}$$

$$P_{116}$$

$$P_{117}$$

$$P_{118}$$

$$P_{119}$$

$$P_{120}$$

$$P_{121}$$

$$P_{122}$$

$$P_{123}$$

$$P_{124}$$

$$P_{125}$$

$$P_{126}$$

$$P_{127}$$

$$P_{128}$$

$$P_{129}$$

$$P_{130}$$

$$P_{131}$$

$$P_{132}$$

$$P_{133}$$

$$P_{134}$$

$$P_{135}$$

$$P_{136}$$

$$P_{137}$$

$$P_{138}$$

$$P_{139}$$

$$P_{140}$$

$$P_{141}$$

$$P_{142}$$

$$P_{143}$$

$$P_{144}$$

$$P_{145}$$

$$P_{146}$$

$$P_{147}$$

$$P_{148}$$

$$P_{149}$$

$$P_{150}$$

$$P_{151}$$

$$P_{152}$$

$$P_{153}$$

$$P_{154}$$

$$P_{155}$$

$$P_{156}$$

$$P_{157}$$

$$P_{158}$$

$$P_{159}$$

$$P_{160}$$

$$P_{161}$$

$$P_{162}$$

$$P_{163}$$

$$P_{164}$$

$$P_{165}$$

$$P_{166}$$

$$P_{167}$$

$$P_{168}$$

$$P_{169}$$

$$P_{170}$$

$$P_{171}$$

$$P_{172}$$

$$P_{173}$$

$$P_{174}$$

$$P_{175}$$

$$P_{176}$$

$$P_{177}$$

$$P_{178}$$

$$P_{179}$$

$$P_{180}$$

$$P_{181}$$

$$P_{182}$$

$$P_{183}$$

$$P_{184}$$

$$P_{185}$$

$$P_{186}$$

$$P_{187}$$

$$P_{188}$$

$$P_{189}$$

$$P_{190}$$

$$P_{191}$$

$$P_{192}$$

$$P_{193}$$

$$P_{194}$$

$$P_{195}$$

$$P_{196}$$

$$P_{197}$$

$$P_{198}$$

$$P_{199}$$

$$P_{200}$$

$$P_{201}$$

$$P_{202}$$

$$P_{203}$$

$$P_{204}$$

$$P_{205}$$

$$P_{206}$$

$$P_{207}$$

$$P_{208}$$

$$P_{209}$$

$$P_{210}$$

$$P_{211}$$

$$P_{212}$$

$$P_{213}$$

$$P_{214}$$

$$P_{215}$$

$$P_{216}$$

$$P_{217}$$

$$P_{218}$$

$$P_{219}$$

$$P_{220}$$

$$P_{221}$$

$$P_{222}$$

$$P_{223}$$

$$P_{224}$$

$$P_{225}$$

$$P_{226}$$

$$P_{227}$$

$$P_{228}$$

$$P_{229}$$

$$P_{230}$$

$$P_{231}$$

$$P_{232}$$

$$P_{233}$$

$$P_{234}$$

$$P_{235}$$

$$P_{236}$$

$$P_{237}$$

$$P_{238}$$

$$P_{239}$$

$$P_{240}$$

$$P_{241}$$

$$P_{242}$$

# Realized Variance

$$\hat{\sigma}_h^2 = \frac{p!}{(n-k)!}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 p(x) dx$$

- Variance measure computed using ultra-high-frequency data (UHF)
  - ▶ Uses all available information to estimate the variance over some period
    - Usually 1 day
  - ▶ Variance estimates from *RV* can be treated as “observable”
    - Standard ARMA modeling
    - Variance estimates are consistent
    - Asymptotically unbiased
    - Variance converges to 0 as the number of samples increases
  - ▶ Problems arise when applied to market data
    - Noise (bid-ask bounce)
    - Market closure
    - Prices discrete
    - Prices not continuously observable
    - Data quality

# Realized Variance

$$\hat{\sigma}_B^2 = \frac{p!}{(n-k)!}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 p(x) dx$$

## ■ Assumptions

- ▶ Log-prices are generated by an arbitrage-free semi-martingale
  - Prices are observable
  - Prices can be sampled often

## ▶ Defined

$$RV_t^{(m)} = \sum_{i=1}^m (p_{i,t} - p_{i-1,t})^2 = \sum_{i=1}^m r_{i,t}^2.$$

- $m$ -sample Realized Variance
- $p_{i,t}$  is the  $i$ th log-price on day  $t$
- $r_{i,t}$  is the  $i$ th return on day  $t$

- ▶ Only uses information on day  $t$  to estimate the variance on day  $t$
- ▶ Consistent estimator of the integrated variance

$$\int_t^{t+1} \sigma_s^2 ds$$

- ▶ “Total variance” on day  $t$

# Why Realized Variance Works



- Consider a simple Brownian motion

$$dp_t = \mu dt + \sigma dW_t$$

- $m$ -sample Realized Variance

$$RV_t^{(m)} = \sum_{i=1}^m r_{i,t}^2$$

- Returns are IID normal

$$r_{i,t} \stackrel{\text{i.i.d.}}{\sim} N\left(\frac{\mu}{m}, \frac{\sigma^2}{m}\right)$$

- Nearly unbiased

$$\mathbb{E}\left[RV_t^{(m)}\right] = \frac{\mu^2}{m} + \sigma^2$$

- Variance close to 0

$$\text{V}\left[RV_t^{(m)}\right] = 4\frac{\mu^2\sigma^2}{m^2} + 2\frac{\sigma^4}{m}$$

# Why Realized Variance Works



- Works for models with time-varying drift and stochastic volatility

$$dp_t = \mu_t dt + \sigma_t dW_t$$

- ▶ No arbitrage imposes some restrictions on  $\mu_t$
- ▶ Works with price processes with jumps
- ▶ In the general case:

$$RV_t^{(m)} \xrightarrow{p} \int_0^1 \sigma_s^2 ds + \sum_{n=1}^N J_n^2$$

- ▶  $J_n$  are jumps

# Why Realized Variance Doesn't Work

- Multiple prices at the same time
  - ▶ Define the price as the average share price (volume weighted price)
  - ▶ Use simple average or median
  - ▶ Not a problem
- Prices only observed on a discrete grid
  - ▶ \$.01 or £.0025
  - ▶ Nothing can be done
  - ▶ Small problem
- Data quality
  - ▶ UHF price data is generally messy
  - ▶ Typos
  - ▶ Wrong time-stamps
  - ▶ Pre-filter to remove obvious errors
  - ▶ Often remove “round trips”
- No price available at some point in time
  - ▶ Use the last observed price: *last price interpolation*
  - ▶ Averaging prices before and after leads to bias

# Solutions to bid-ask bounce type noise

- Bid-ask bounce is a **big** problem

- ▶ Simple model with “pure” noise

$$p_{i,t} = p_{i,t}^* + \nu_{i,t}$$

- $p_{i,t}$  is the observed price with noise
    - $p_{i,t}^*$  is the unobserved efficient price
    - $\nu_{i,t}$  is the noise

- ▶ Easy to show

$$r_{i,t} = r_{i,t}^* + \eta_{i,t}$$

- $r_{i,t}^*$  is the unobserved efficient return
    - $\eta_{i,t} = \nu_{i,t} - \nu_{i-1,t}$  is a MA(1) error

- ▶  $RV$  is badly biased

$$RV_t^{(m)} \approx \widehat{RV}_t + m\tau^2$$

- Bias is increasing in  $m$
    - Variance is also increasing in  $m$

# Simple solution

$$\hat{\sigma}_B^2 = \frac{\mu!}{(\mu - k)!}$$



$$\Omega_B = \int_{-\infty}^{+\infty} (x - M_B)^2 p(x) dx$$

- Do not sample frequently

- ▶ 5-30 minutes
  - Better than daily but still inefficient
- ▶ Remove MA(1) by filtering
  - $\eta_{i,t}$  is an MA(1)
  - Fit an MA(1) to observed returns

$$r_{i,t} = \theta \epsilon_{i-1,t} + \epsilon_{i,t}$$

- Use fit residuals  $\hat{\epsilon}_{i,t}$  to compute  $RV$ 
  - Generally biased downward
- ▶ Use mid-quotes
  - A little noise
  - My usual solution

# A modified Realized Variance estimator: $RV^{AC1}$

- Best solution is to use a modified  $RV$  estimator

- ▶  $RV^{AC1}$

$$RV_t^{AC1(m)} = \sum_{i=1}^m r_{i,t}^2 + 2 \sum_{i=2}^m r_{i,t} r_{i-1,t}$$

- ▶ Adds a term to  $RV$  to capture the MA(1) noise
- ▶ Looks like a simple Newey-West estimator
- ▶ Unbiased in pure noise model
- ▶ Not consistent
- ▶ Realized Kernel Estimator
  - Adds more weighted cross-products
  - Consistent in the presence of many realistic noise processes
  - Fairly easy to implement

# One final problem

$$\hat{\sigma}_B^2 = \frac{\mu!}{(\mu - k)!}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

## ■ Market closure

- ▶ Markets do not operate 24 hours a day (in general)
- ▶ Add in close-to-open return squared

$$RV_t^{(m)} = r_{\text{CtO},t}^2 + \sum_{i=1}^m r_{i,t}^2$$

—  $r_{\text{CtO},t} = p_{\text{Open},t} - p_{\text{Close},t-1}$

- ▶ Compute a modified  $RV$  by weighting the overnight and open hour estimates differently

$$\widetilde{RV}_t^{(m)} = \lambda_1 r_{\text{CtO},t}^2 + \lambda_2 RV_t^{(m)}$$

# The volatility signature plot



- Hard to know how often to sample
  - ▶ Visual inspection may be useful

## Definition (Volatility Signature Plot)

The volatility signature plot displays the time-series average of Realized Variance

$$\overline{RV}_t^{(m)} = T^{-1} \sum_{t=1}^T RV_t^{(m)}$$

as a function of the number of samples,  $m$ . An equivalent representation displays the amount of time, whether in calendar time or tick time (number of trades between observations) along the X-axis.

# Some empirical results

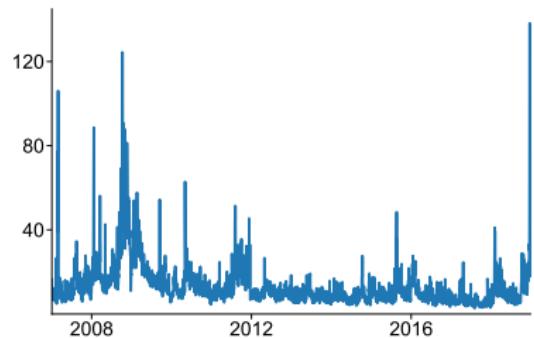


$$\Omega_x = \int_{-\infty}^{\infty} (x - M_x)^2 \phi(x) dx$$

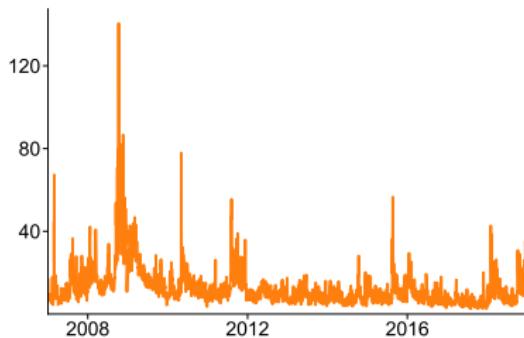
- S&P 500 Depository Receipts
  - ▶ SPiDeRs
  - ▶ AMEX: SPY
  - ▶ Exchange Traded Fund
  - ▶ Ultra-liquid
    - 100M shares per day
    - Over 100,000 trades per day
    - 23,400 seconds in a typical trading day
  - ▶ January 1, 2007 – December 31, 2018
  - ▶ Filtered by daily High-Low data
  - ▶ Some cleaning of outliers

# SPDR Realized Variance ( $RV$ )

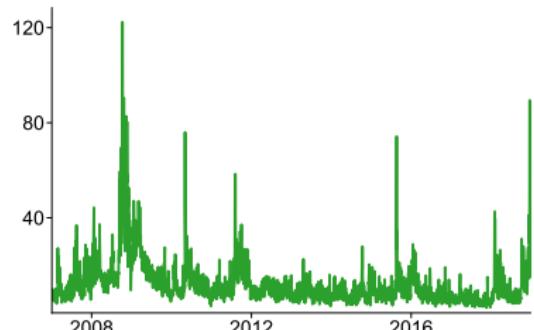
$RV$ , 15 seconds



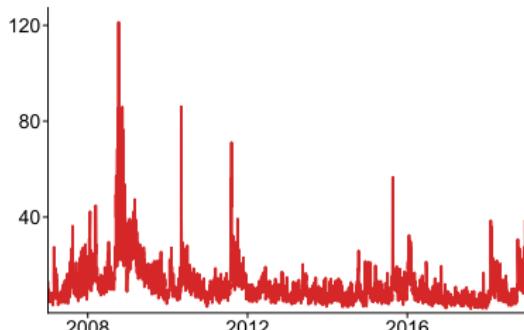
$RV$ , 1 minute



$RV$ , 5 minutes

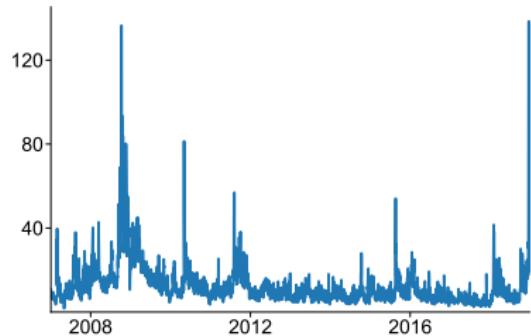


$RV$ , 15 minutes

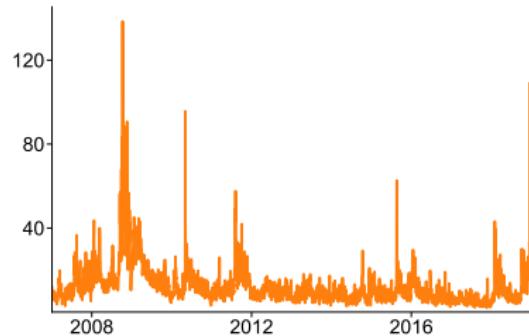


# SPDR Realized Variance ( $RV^{AC1}$ )

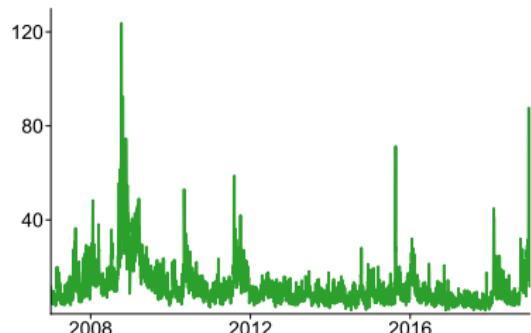
$RV^{AC1}$ , 15 seconds



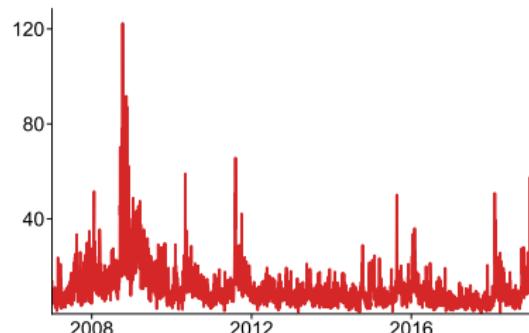
$RV^{AC1}$ , 1 minute



$RV^{AC1}$ , 5 minutes

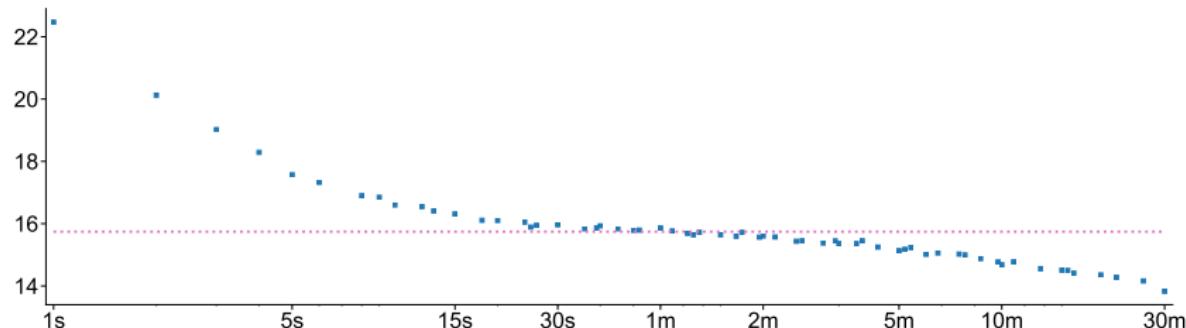


$RV^{AC1}$ , 15 minutes

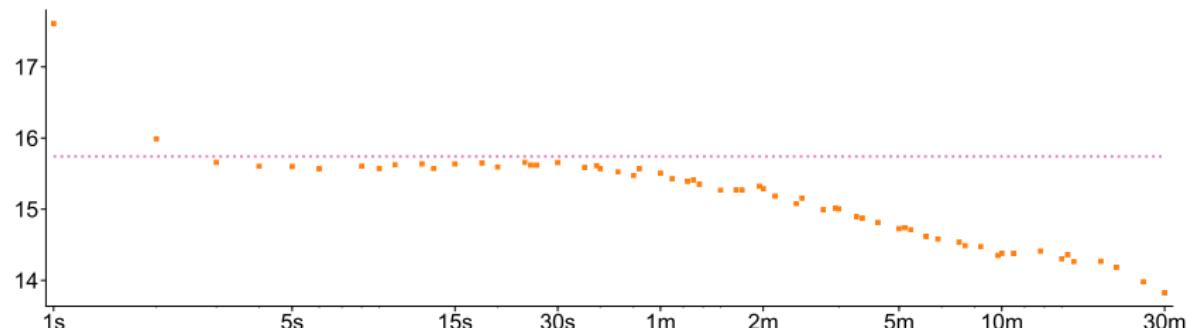


# Volatility Signature Plots

Volatility Signature Plot for SPDR  $RV$

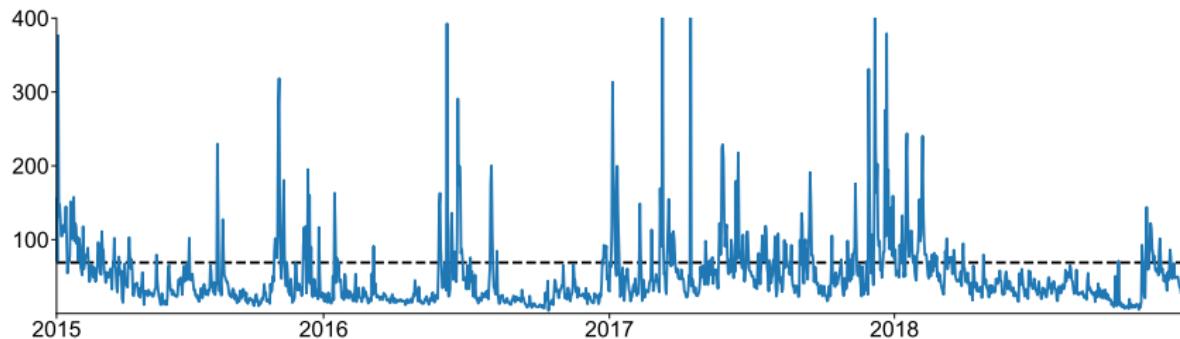


Volatility Signature Plot for SPDR  $RV^{AC1}$

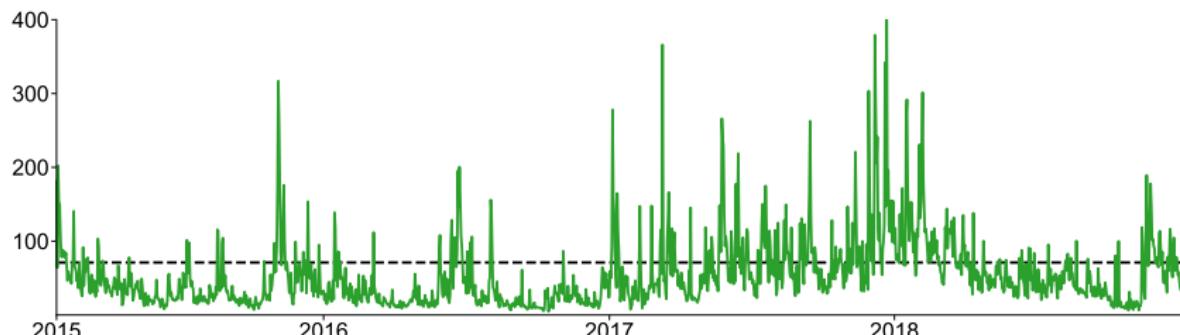


# Bitcoin Realized Variance

5-second  $RV$



5-minute  $RV$



# Modeling Realized Variance



$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - M_x)^2 \phi(x) dx$$

- Two choices
  - Treat volatility as observable and model as ARMA
    - ▶ Really simply to do
    - ▶ Forecasts are equally simple
    - ▶ Theoretical motivation why RV may be well modeled by an ARMA( $P, 1$ )
  - Continue to treat volatility as latent and use ARCH-type model
    - ▶ Realized Variance is still measured with error
    - ▶ A more precise measure of conditional variance than daily returns squared,  $r_t^2$ , but otherwise similar

# Treating $\sigma_t^2$ as observable



$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - M_x)^2 w(x) dx$$

- If  $RV$  is  $\sigma_t^2$ , then variance is observable
- Main model used is a Heterogeneous Autoregression
- Restricted AR(22) in levels

$$RV_t = \phi_0 + \phi_1 RV_{t-1} + \phi_5 \overline{RV}_{5,t-1} + \phi_{22} \overline{RV}_{22,t-1} + \epsilon_t$$

- Or in logs

$$\ln RV_t = \phi_0 + \phi_1 \ln RV_{t-1} + \phi_5 \ln \overline{RV}_{5,t-1} + \phi_{22} \ln \overline{RV}_{22,t-1} + \epsilon_t$$

where  $\overline{RV}_{j,t-1} = j^{-1} \sum_{i=1}^j RV_{t-i}$  is a  $j$  lag moving average

- Model picks up volatility changes at the daily, weekly, and monthly scale
- Fits and forecasts RV fairly well
  - ▶ MA term may still be needed

# Leaving $\sigma_t^2$ latent

$$\hat{\sigma}_t^2 = \frac{p!}{(n-k)!}$$



$$\Omega_x = \int_{-\infty}^{+\infty} (x - M_x)^2 w(x) dx$$

- Alternative if to treat RV as a proxy of the latent variance and use a *non-negative multiplicative error model* (MEM)
- MEMs specify the mean of a process as  $\mu_t \times \psi_t$  where  $\psi_t$  is a mean 1 shock.
- A  $\chi_1^2$  is a natural choice here
- ARCH models are special cases of a non-negative MEM model
- Easy to model RV using existing ARCH models
  - 1. Construct  $\tilde{r}_t = \text{sign}(r_t) \sqrt{RV_t}$
  - 2. Use standard ARCH model building to construct a model for  $\tilde{r}_t$

$$\sigma_t^2 = \omega + \alpha_1 \tilde{r}_{t-1}^2 + \gamma_1 \tilde{r}_{t-1}^2 I_{[\tilde{r}_{t-1} < 0]} + \beta_1 \sigma_{t-1}^2$$

becomes

$$\sigma_t^2 = \omega + \alpha_1 RV_{t-1} + \gamma_1 RV_{t-1} I_{[r_{t-1} < 0]} + \beta_1 \sigma_{t-1}^2$$

$$n! \approx \left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}$$

$$A_n^k = \frac{n!}{(n-k)!}$$

$$\rho_n = \frac{n!}{(n-n)!} = \frac{n!}{0!}$$

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)$$

$$\lambda_n^k = n \cdot (n-1) \cdot \dots \cdot n = n^k$$

$$\tilde{\rho}_{n_1, n_2, \dots, n_k} = \frac{(n_1 + n_2 + \dots + n_k)!}{n_1! n_2! \dots n_k!}$$

$$C_n^k = \frac{n!}{k!(n-k)!}$$

$$\tilde{C}_n^m = \frac{(n+m-1)!}{m!(n-1)!}$$

$$(a+b)^n = C_p^0 a^p + C_p^1 a^{p-1} b^1 + \dots + C_p^{p-1} a^1 b^{p-1} + C_p^p b^n = \sum_{k=0}^n C_p^k a^{p-k} b^k$$

$$p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + p(B|A_3)p(A_3) + \dots + p(B|A_k)p(A_k)$$

$$\rho(x) = \frac{p(B|A_1)p(A_1)}{p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \dots + p(B|A_k)p(A_k)}$$

$$P_{\mu}(N)$$

# Implied Variance

$$\sigma_{\text{obs}}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$D_x = \hat{\omega}^2 = M_x^2 - (M_x)^2$$

$$\rho_{\varepsilon}(z) = \frac{\lambda^z}{\rho^z} e^{-\lambda}$$

$$p(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \phi(x) dx$$

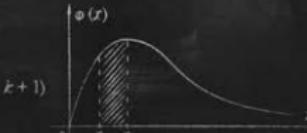
$$M_x = \sum_{i=1}^k p_i x_i$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p = \lim_{N \rightarrow \infty} \frac{f_i}{N}$$

$$C = \frac{PE_n S}{d}$$

$$D_x = \sum_{i=1}^k p_i (x_i - M_x)^2$$



$$D_x = \int_{-\infty}^{+\infty} (x - M_x)^2 \phi(x) dx$$

$$M_x = \int_{-\infty}^{+\infty} x \cdot \phi(x) dx$$

$$V_{f(x)} = \int_{-\infty}^{+\infty} f(x) \phi(x) dx$$

$$S = \eta_0^2 c + \frac{m^2}{2}$$

$$F = G \frac{m_1 m_2}{r^2}$$

$$f(x) = \delta g \left( \frac{x_0}{2\pi k T} \right)^N e^{-\frac{m_0^2}{2\pi k T}}$$

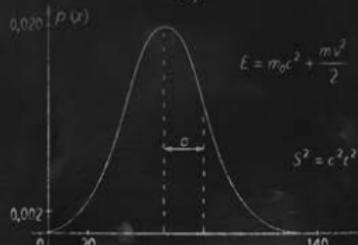


$$\vec{d} = \frac{m_0}{2\pi k T} (\cos \alpha_1 - \cos \alpha_2)$$



$$d^2 = A_1^2 + A_2^2 + 2 A_1 A_2 \cos(\phi_2 - \phi_1)$$

$$\hbar v = A + \frac{m v^2}{2}$$



$$E = m_0 c^2 + \frac{m v^2}{2}$$

$$m = m_0 / \sqrt{1 - \beta^2}$$

$$S^2 = c^2 t^2 - l^2 = i \hbar v$$

$$r_n = \frac{4\pi \epsilon_0 n^2 n^2}{m Z e^4}$$

# Implied Volatility and VIX



- Implied volatility is very different from ARCH and Realized measures
- Market based: Level of volatility is calculated from options prices
- Forward looking: Options depend on future price path
- “Classic” implied relies on the Black-Scholes pricing formula
- “Model free” implied volatility exploits a relationship between the second derivative of the call price with respect to the strike and the risk neutral measure
- VIX is a Chicago Board Options Exchange (CBOE) index based on a model free measure
- Allows volatility to be directly traded

# Black-Scholes Implied Volatility

- Black-Scholes Options Pricing
- Prices follow a geometric Brownian Motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

- Constant drift and volatility
- Price of a call is

$$C(T, K) = S\Phi(d_1) + Ke^{-rT}\Phi(d_2)$$

where

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2) T}{\sigma \sqrt{T}}$$

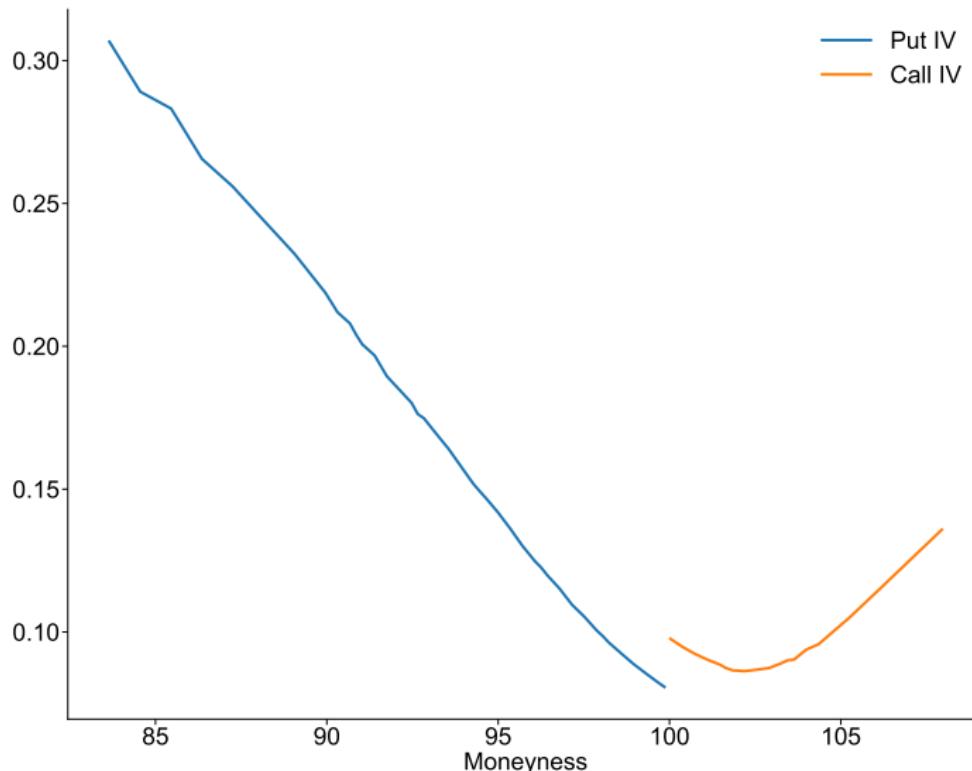
$$d_2 = \frac{\ln(S/K) + (r - \sigma^2/2) T}{\sigma \sqrt{T}}.$$

- Can invert to produce a formula for the volatility given the call price  $C(T, K)$

$$\sigma_t^{\text{Implied}} = g(C_t(T, K), S_t, K, T, r)$$

# BSIV against Moneyness for SPY

Jan 15, 2018 options expiring on Feb 2, 2018



# Model Free Implied Volatility



- Model free uses the relationship between option prices and RN density
- The price of a call option with strike  $K$  and maturity  $t$  is

$$C(t, K) = \int_K^{\infty} (S_t - K) \phi_t(S_t) dS_t$$

- $\phi_t(S_t)$  is the *risk-neutral* density at maturity  $t$
- Differentiating with respect to strike yields

$$\frac{\partial C(t, K)}{\partial K} = - \int_K^{\infty} \phi_t(S_t) dS_t$$

- Differentiating again with respect to strike yields

$$\frac{\partial^2 C(t, K)}{\partial K^2} = \phi_t(K)$$

- The change in an option price as a function of the strike  $K$  is the probability of the stock price having value  $K$  at time  $t$
- Allows for risk-neutral density to be recovered from a continuum of options *without assuming a model for stock prices*

# Model Free Implied Volatility


$$\sigma_x = \sqrt{\int_0^{M_x} (x - M_x)^2 \pi(x) dx}$$

- The previous result allows a model free IV to be computed from

$$E_{\mathbb{F}} \left[ \int_0^t \left( \frac{\partial F_s}{F_s} \right)^2 ds \right] = 2 \int_0^\infty \frac{C^F(t, K) - (F_0 - K)^+}{K^2} dK$$

- Devil is in the details

- Only finitely many calls
- Thin trading
- Truncation

$$\sum_{m=1}^M [g(T, K_m) + g(T, K_{m-1})] (K_m - K_{m-1})$$

where

$$g(T, K) = \frac{C(t, K/B(0, t)) - (S_0 - K)^+}{K^2}$$

- See Jiang & Tian (2005, *RFS*) for a very useful discussion

$$\mu \approx \left(\frac{\sigma}{\delta}\right)^2 \sqrt{2\ln 2}$$

$$\hat{\rho}_B^T = \frac{\mu}{(\mu - k)}$$



$$\Omega_x := \int_0^{\infty} (x + M_x)^{-1} p(x) dx$$

- VIX is continuously computed by the CBOE
- Uses a model-free style formula
- Uses both calls and puts
- Focuses on out-of-the-money options
  - ▶ OOM options are more liquid
- Formula:

$$\sigma^2 = \frac{2}{T} e^{rT} \sum_{i=1}^N \frac{\Delta K_i}{K_i^2} Q(K_i) - \frac{1}{T} \left( \frac{F_0}{K_0} - 1 \right)^2$$

- ▶  $Q(K_i)$  is the mid-quote for a strike of  $K_i$ ,  $K_0$  is the first strike below the forward index level
- ▶ Only uses out-of-the-money options
- ▶ VIX appears to have information about future *realized volatility* that is not in other backward looking measures (GARCH/RV)

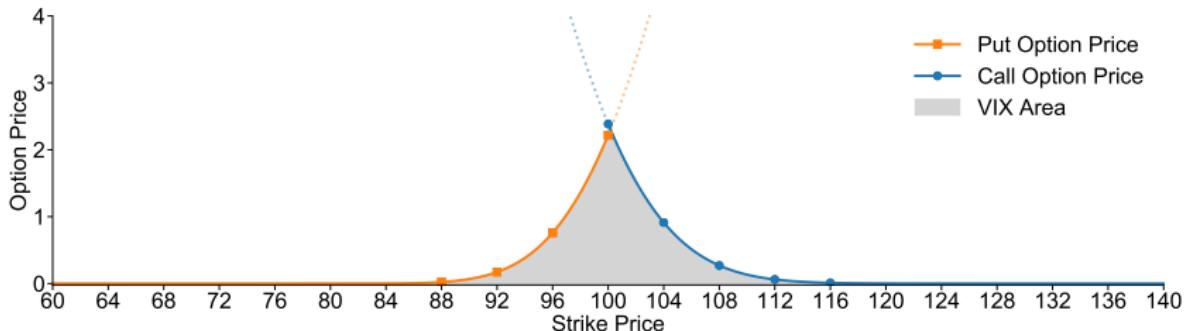
# Model-Free Example

- MFIV works under weak conditions on the underlying price process
  - ▶ Geometric Brownian motion is included
- Put and call options prices computed from Black-Scholes
  - ▶ Annualized volatility either 20% or 60%
  - ▶ Risk-free rate 2%, time-to-maturity 1 month ( $T = 1/12$ )
  - ▶ Current price 100 (normalized to moneyness), strikes every 4%
- Contribution is  $\frac{2}{T} e^{rT} \frac{\Delta K_i}{K_i^2} Q(K_i)$

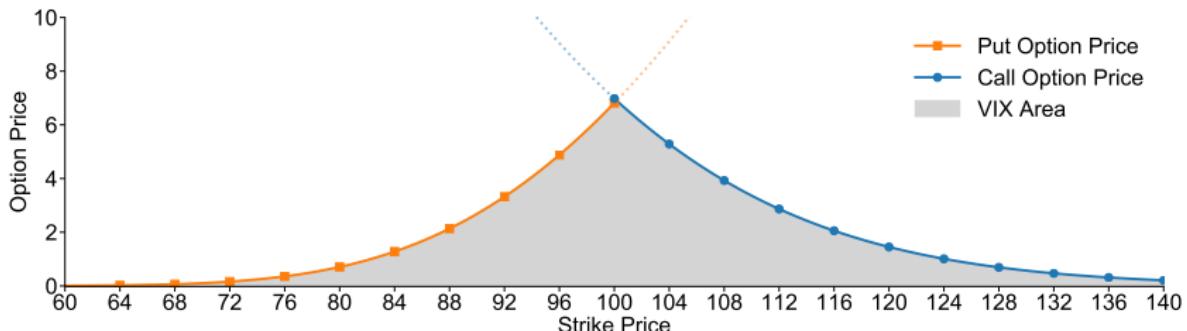
Strike	Call	Put	Abs. Diff.	VIX Contrib.
88	12.17	0.02	12.15	0.0002483
92	8.33	0.17	8.15	0.0019314
96	4.92	0.76	4.16	0.0079299
100	2.39	2.22	0.17	0.0221168
104	0.91	4.74	3.83	0.0080904
108	0.27	8.09	7.82	0.0022259
112	0.06	11.88	11.81	0.0004599
116	0.01	15.82	15.81	7.146e-05
Total				0.0430742

# Model-Free Example

## 20% Annualized Volatility

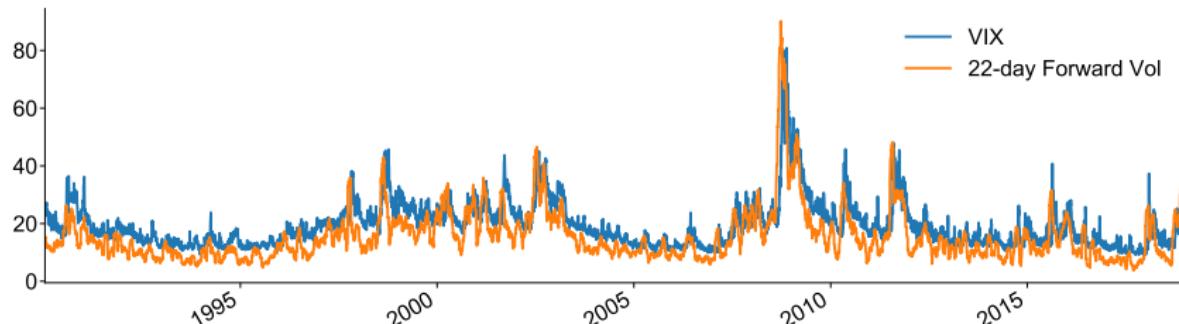


## 60% Annualized Volatility

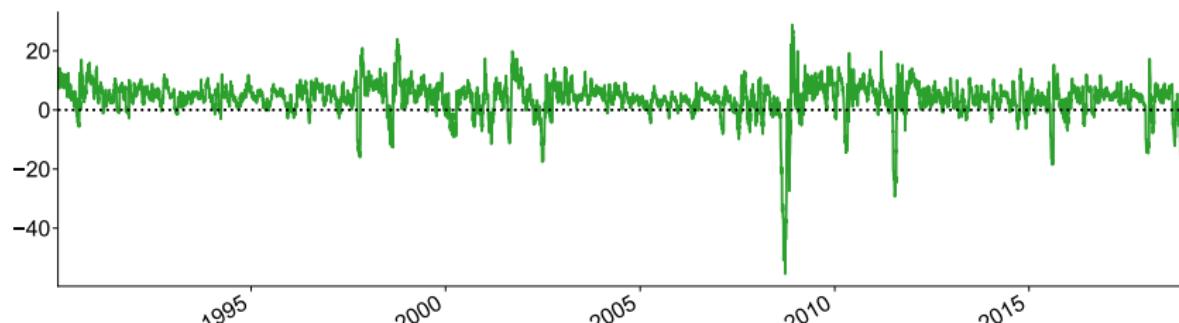


# VIX against TARCH(1,1,1) Forward-vol

## VIX and Forward Volatility



## VIX Forward Volatility Difference



# Variance Risk Premium

- Difference between VIX and forward volatility is a measure of the return to selling volatility
- Variance Risk Premium is strictly forward looking

$$E_t^{\mathbb{Q}} \left[ \int_0^{t+h} \left( \frac{\partial F_s}{F_s} \right)^2 ds \right] - E_t^{\mathbb{P}} \left[ \int_t^{t+h} \left( \frac{\partial F_s}{F_s} \right)^2 ds \right]$$

- Defined as the difference between RN ( $E^{\mathbb{Q}}$ ) and physical ( $E^{\mathbb{P}}$ ) variance
  - ▶ RN variance measured using VIX or other MFIV
  - ▶ Physical forecast from HAR or other model based on Realized Variance
    - RV matters, using daily is sufficiently noisy that prediction is not useful









$$E[c_+^4 \alpha_+] = E[\alpha_+^4 E_{\pm-} [c_+]]$$

$c_+ \sim N(0, 1)$

$$= 3 E[\alpha_+^4]$$











